Rational altruism? On preference estimation and dictator game experiments

Grech, Philip D ; Nax, Heinrich H

Abstract: Experimental implementations of dictator games are found to differ in terms of their underlying strategic incentives. We explore this discovery in two separate directions. Theoretically, assuming identical other-regarding preferences, we show that the two most widely used protocols can generate strongly contrasting rational-choice predictions, from which different interpretations of dictator giving arise. Experimentally, a tailor-made experiment reveals significant differences between the two protocols but rejects full rationality as a satisfactory explanatory theory. Our findings indicate that several previously drawn conclusions regarding other-regarding preferences among humans distinguished by social class, gender, generation, nationality, etc. may be more ambiguous than hitherto believed.

DOI: https://doi.org/10.1016/j.geb.2019.10.004

Games and Economic Behavior 119 (2020) 309–338

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Rational altruism? On preference estimation and dictator game experiments

Philip D. Grech [a,*], Heinrich H. Nax [b,c,*]

[a] Department of Management, Technology and Economics, ETH Zurich, Scheuchzerstrasse 7, 8092 Zurich, Switzerland
[b] Behavioral Game Theory, ETH Zurich, Clausiusstrasse 37, 8092 Zurich, Switzerland
[c] Institute of Sociology, University of Zurich, Andreasstrasse 15, 8050 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 1 June 2017
Available online 22 October 2019

JEL classification:
A13
C72
D01
D64

Keywords:
Altruism
Charitable giving
Dictator games
CES utility functions
Distributional preferences
Social preferences
Experimental economics
Foundations

ABSTRACT

Experimental implementations of dictator games are found to differ in terms of their underlying strategic incentives. We explore this discovery in two separate directions. Theoretically, assuming identical other-regarding preferences, we show that the two most widely used protocols can generate strongly contrasting rational-choice predictions, from which different interpretations of dictator giving arise. Experimentally, a tailor-made experiment reveals significant differences between the two protocols but rejects full rationality as a satisfactory explanatory theory. Our findings indicate that several previously drawn conclusions regarding other-regarding preferences among humans distinguished by social class, gender, generation, nationality, etc. may be more ambiguous than hitherto believed.

'We take license of calling the dictator game a "game" although it is a single person decision problem.'
Forsythe et al. (1994)

1. Introduction

In the original formulation of the dictator game (Kahneman et al., 1986; Forsythe et al., 1994), one person acting in the role of a 'dictator' decides how to split a pie of fixed size, or, equivalently, how much of his own wealth to redistribute to someone else at a one-to-one exchange rate; see Fig. 1a for a stylized illustration. Understanding how individuals behave in such situations and why they do so has important implications for explaining human behavior in general, and, in particular, for a wide array of giving-contexts such as cooperating, donating, negotiating, helping, etc.

From the perspective of narrow self-interest, a dictator maximizes his material payoff by keeping everything and giving nothing. The fact that positive giving is consistently observed in controlled laboratory experiments ever since Kahneman

* Corresponding authors.
E-mail addresses: pgrech@ethz.ch (P.D. Grech), heinrich.nax@uzh.ch (H.H. Nax).

https://doi.org/10.1016/j.geb.2019.10.004

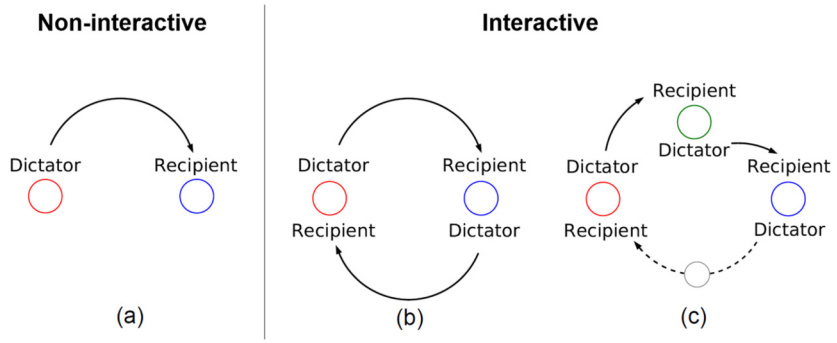

Fig. 1. Types of Dictator Games. (*a*) *Non-interactive* (two players, one dictator): the dictator's own material payoff depends only on how much he/she keeps, and the recipient's material payoff depends only on how much the dictator gives. (*b*) *Interactive* (two players, two dictators): each player is both, the other player's dictator and recipient. (*c*) *Interactive* (*n* players, *n* dictators): each player is someone's dictator and someone *else*'s recipient, etc. - until the 'loop' closes.

et al. (1986) falsifies the hypothesis that all humans are always narrowly motivated by material self-interest.[1] In order to move on and measure more precisely how other-regarding (also referred to as distributional, social, altruistic etc.) concerns matter beyond narrow self-interest, generalized dictator game experiments have been conducted where the multiplier of redistribution is varied meaning that a dictator's giving may have different worth in the recipient's hands. Giving decisions across a range of thus modified dictator games have been used to calibrate various other-regarding preference types.

Inferring preferences from giving decisions in this way relies on the fundamental assumption that individual actions accurately 'reveal' preferences in the sense that all payoff-relevant factors (including beliefs) are expressed (Samuelson, 1938). Andreoni and Miller (2002) made a seminal contribution translating methods based on the generalized axiom of revealed preferences (GARP) from consumer theory to dictator games, thus proposing a framework to estimate other-regarding preferences based on classical rationality axioms (hence 'rational altruism'). The key insight of this literature is to interpret individual giving as rational decisions driven by individual other-regarding preferences, an approach sometimes subsumed under the umbrella of the 'subjective expected utility correction project' (Gigerenzer and Selten, 2001). Andreoni and Miller (2002)'s method is widely used (see e.g. Fisman et al. 2007, 2015b,a), prominently to fit utility functions of the constant-elasticity-of-substitution (CES) family which allow classifications of individuals according to their trade-offs of self-interest vs. altruism and of equality vs. efficiency; see Jakiela (2013) for a literature overview.[2] Related methods based on different functional assumptions regarding utilities include, for example, the classic 'ring measure' from social psychology (Liebrand, 1984; Murphy et al., 2011; Nax et al., 2015), and other other-regarding preferences such as those due to Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Charness and Rabin (2002).

All other-regarding preference estimation techniques, whether based on CES utilities or on other preference models, rely crucially on a special feature of the dictator game as it was originally formulated (henceforth referred to as the 'non-interactive' dictator game), which is that giving decisions are not interdependent: the dictator is no one's recipient and the recipient is no one's dictator (as illustrated in Fig. 1a). If the decision context is instead strategic, as is the case when a player gives and receives at the same time (thus yielding a truly interdependent one-shot game as illustrated in Figs. 1b & 1c), then giving decisions cannot be disentangled from beliefs regarding others' giving decisions, and individual preferences cannot be inferred in the same way. In such 'interactive' dictator games where each player assumes both roles –dictator and recipient– at the same time, each player's total monetary payoff is, on the one hand, determined by the payment he/she receives from 'his/her' dictator, and, on the other hand, by the amount he/she keeps and does not give to 'his/her' recipient. Thus, the payoff to any given player is a function of his/her own decision *and* those of others.[3]

The key motivation for the present paper was our discovery that many dictator game studies that aim to measure preferences have been and continue to be implemented interactively, but are then analyzed within the framework of non-interactive dictator games.[4] To the best of our knowledge, the implications of such protocol pooling have so far not undergone detailed scrutiny in the scholarly literature.[5] This is also illustrated by Engel's extensive (2011) meta-study of the dictator game experiments literature which does not differentiate along this dimension. Similarly, Jakiela (2013)'s review of dictator game studies that focuses on measuring other-regarding preferences does not mention this issue either.

[1] See Engel (2011) for an extensive meta-study involving 131 papers.

[2] GARP goes back to Samuelson (1938), CES functions to Solow (1956), and preference estimation to Afriat (1967, 1972); see Choi et al. (2007), Polisson and Quah (2013), Polisson and Renou (2016) for extensions.

[3] Interactive dictator games constitute interesting games in their own right, as they can be used to model situations such as gift exchange or mutual help, where the gift one gives or the help one offers to others may explicitly depend on the gift or the help one receives.

[4] In fact, the list of interactive implementations doing this includes Andreoni and Miller (2002) itself. Other prominent examples include studies that draw conclusions regarding differences in pro-sociality between humans distinguished by factors such as social class (Fisman et al., 2015b), gender (Andreoni and Vesterlund, 2001), generation (Cameron et al., 2013), nationality (Ashraf et al., 2006), and more.

[5] We are only aware of two experimental results, Greiff et al. (2018), Korenok et al. (2013), which will be discussed below.

However, treating non-interactive and interactive dictator games the same has consequences regarding the interpretation of such experiments, in particular regarding the rationality assumptions that pertain to their analysis. While rational-choice assumptions allow for other-regarding preference estimation as outlined above under non-interactive protocols, the applicability of such techniques for interactive protocols relies on a rather extreme view of rationality. Concretely, it requires that the economic agent is, on the one hand, perfectly rational in terms of pursuing utility-maximization that takes into account the welfare of others, while, on the other hand, he/she is perfectly irrational in a strategic sense and unaware that others take decisions that are relevant for himself/herself. That is, he/she is highly rational in terms of his/her effect on others' welfare, but entirely unable of any 'cognitive empathy' in terms of putting himself/herself in others' shoes.

If maintained, the kind of strategic unawareness outlined above would either be a natural trait of humans or it would have been created by the experimenter, namely if subjects were successfully framed so as to perceive an interactive setting as a non-interactive one. The former case conflicts with findings from many other –often more sophisticated– experimental games, the latter might violate important principles of experimental economics. Indeed, following Bardsley et al. (2010), subjects should be provided with instructions that bring across accurately the true nature of the underlying decision context: a comparison with (theoretical) rational-choice benchmarks of an experimental game becomes meaningful only when subjects can reasonably be assumed to have a correct interpretation of the underlying strategic structure. Thus, either all payoff consequences of actions ought to be explained to subjects, or, whatever is not being explained ought to be made explicit. And indeed, most studies in experimental game theory rely on instructions where a great deal of effort is made to make strategic interdependencies explicit (by using matrices, figures, examples, etc., for example, in voluntary contributions games, trust games, etc.). All interactive dictator game instructions we are aware of state the interactive nature of the game explicitly (with a sentence or two), but without providing further details to make it particularly easy to understand (controls in our experiment, which is based on standard instructions, actually reveal substantial rates of misunderstanding in all treatments, cf. Footnote 36 below). As a result, we as analysts can neither be sure that subjects understand the interdependencies perfectly (and do not play the game as if it was non-interactive), nor that they do not understand them at all (and do play as if the game was non-interactive).

If, by contrast, we let go of the assumption of strategic unawareness, the two protocols can no longer be used interchangeably, as one of them is non-strategic and the other represents a proper (one-shot) game.[6] Thus, beyond preferences for selfishness and efficiency, differing degrees of strategic sophistication and awareness then influence the observed giving patterns in interactive settings and provide additional confounds. Related to this, Dufwenberg et al. (2011) have shown that a similar interpretational ambiguity arises in competitive equilibrium according to which it is generally impossible to infer other-regarding preferences from behavior.

The goals of the present paper are twofold. First, we develop standard neoclassical rational-choice benchmarks for an interactive dictator game model and compare them with rational giving decisions in non-interactive games.[7] This will constitute an analysis of the interactive dictator games' Nash equilibria where players are characterized by the same kinds of other-regarding preferences that are used in non-interactive dictator games. Second, we conduct a tailor-made experiment in order (*a*) to test for protocol differences between interactive and non-interactive implementations, and, if there are differences, (*b*) to investigate whether rationality benchmarks can explain observed behavior.

Our theory results summarize as follows. In the non-interactive case, rational giving with other-regarding preferences predicts that the amount allocated by a dictator to his/her recipient is an intermediate one, unless the individual is narrowly self-interested (perfectly altruistic), in which case he/she gives zero (everything). In the interactive case, the standard rational-choice benchmark is Nash equilibrium, which accounts for the fact that giving decisions are interdependent. We show that there exist many circumstances under which equilibria are characterized by a bang-bang structure: specifically, if players have sufficiently low (high) concerns for others' payoffs, then zero (full) payments are made. This is true for arbitrary assumptions regarding 'bracketing' (i.e. how weights are being attached to other players), for (some) incomplete information regarding other players' preferences and experimental parameters, and for an array of alternative other-regarding preference specifications (in particular, CES utilities and preference models by Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Charness and Rabin (2002)). Finally, we also show that, under various natural assumptions on players' bracketing and concerns, extremal payments constitute the unique pure-strategy Nash equilibria. These findings enable us to make a clear case in point regarding the inferential problem addressed above: zero giving in interactive dictator games *cannot* unambiguously be interpreted as extreme selfishness as it may also be strategic play by a substantially altruistic individual. Thus, a population sample making zero payments under an interactive design could in principle even be *more* altruistic than another population sample making non-zero payments under a non-interactive protocol. For full disclosure, we point out that our specific equilibrium results may be more or less generic depending on the specific interactive protocol being used – and we wish to abstain from making any claims other than what we can prove. However, we hope that even the skeptical reader can, by virtue of our arguments, appreciate the fact that non-interactive and interactive protocols yield different rational-choice benchmarks – and in many cases even quite strikingly so.

[6] Hypothetically, if, for any given player, all equilibrium payments in the interactive setting were always exactly identical to his/her decisions in the non-interactive setting, the two protocols might be used interchangeably, at least theoretically (less so practically given that subjects are usually not perfectly rational). However, we shall see that interactive dictator games generically feature no dominant strategies.

[7] 'Neoclassical' here simply refers to a utilitarian notion of rationality (allowing other-regarding preferences), see Binmore (2015).

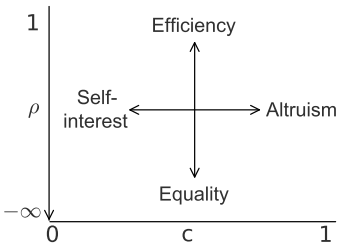

Fig. 2. Space of CES Utilities. Combinations expressing various other-regarding preferences with respect to the tradeoffs of self-interest versus altruism (c on the x-axis) and of equality versus efficiency (ρ on the y-axis).

An experimental investigation of protocol differences complements our theory, providing the basis for additional tests. To follow as methodologically clean a test procedure as possible, we pre-committed a complete analysis plan at the Open Science Framework (OSF) based on a pre-registered design akin to randomized controlled trials as are standardly used in various other disciplines. Our experimental findings summarize as follows. First, we test for protocol differences, and find that significant differences exist. This implies that the differing strategic incentives have *some* –to be understood– effect. Second, we derive a set of testable hypotheses from our theoretical results, and by rejecting these establish that standard rational-choice assumptions cannot organize the data, even though some of the differences bear traces of interactive play.[8]

In combination, both our theoretical and experimental results highlight that the determinants of voluntary giving are more complex than hitherto assumed and suggest that experimental studies relying on interactive protocols are difficult to interpret: taking their (non-interactive) theory part as a given, different experiments seem to be in order, taking their (interactive) experimental part as a given, a different theoretical treatment is warranted.[9]

The remainder of this paper is structured as follows. Next, we outline the models and present our theory results. In the subsequent results sections, we discuss their relevance for dictator game implementations, analyze our own experiment and compare our findings with existing data from the literature that uses both non-interactive and interactive implementations. Finally, summary and outlook conclude.

2. Model

In this section we introduce utility functions with constant elasticity of substitution (CES) representing a wide array of distributional preferences, and illustrate how they can be used to model both non-interactive as well as interactive dictator games. While we shall work with these specific functional assumptions throughout the main text for the sake of presentation, we show that our main results also hold for alternative utility specifications, as will be indicated where appropriate (proofs are relegated to Appendix A).

2.1. CES utility functions

CES preferences are frequently used to explain dictator game decisions, and may be written as

$$v(o,s) := \left[co^{\rho} + (1-c)s^{\rho}\right]^{\frac{1}{\rho}}, \tag{1}$$

where $c \in [0,1]$ and $\rho \in (-\infty, 1]$ are two parameters, whose significance will be clarified below.[10] For $\rho < 0$, if $s = 0$ or $o = 0$ we define v by left-continuous extension, i.e. $v(s,0) = v(0,o) = v(0,0) = 0$. If $\rho = 0$ then v is defined as a Cobb-Douglas utility function $v(o,s) = o^c s^{1-c}$, which, by the Bernoulli-de-l'Hôpital rule, arises as the pointwise limit of (1) if $c \neq 0, 1$. The terms $o, s \geq 0$ denote the final material 'payoffs' of *self* (in the dictating role) and *other* (in the receiving role), respectively.

Fig. 2 illustrates the parameter range of CES utility functions. To provide some intuition for such preferences and how they depend on parameters c and ρ, we consider the following limiting cases.

Perfect Egoist. If $c = 0$, then $v(o,s) = s$: the player only seeks to maximize his own payoff.
Perfect Altruist. If $c = 1$, however, $v(o,s) = o$: the player maximizes the recipients payoff.
Rawlsian. If $\rho \to -\infty$ and $c \neq 0, 1$ we obtain Leontief utility functions $v(o,s) = \min\{o,s\}$ with L-shaped indifference curves: the player seeks equality.

[8] The fact that the game theorists who commented on our work thus far were as surprised by our theoretical results as we were, is perhaps an indicator that Nash equilibrium predictions are unlikely to be observed when tested with experimental subjects – as is the case with many other one-shot experimental games.

[9] Results of past studies may still carry truth, in particular if we knew that subjects were strategically unaware – an assumption which is however troublesome in its own right as was laid out above.

[10] See Jakiela (2013) for a detailed discussion of characteristics and usages of CES preferences.

Efficiency optimizer. If $\rho = 1$ and $c \neq 0, 1$ the utility function becomes linear, $v(o, s) = co + (1 - c)s$, so that, in presence of a budget constraint, the player is purely efficiency-oriented.

Thus, c measures the player's concern and guides from self-interest to altruism, while ρ quantifies the player's efficiency orientation ranging from equality-concern to pure efficiency-orientation.

2.2. Non-interactive dictator games

In the standard non-interactive setting, a single dictator possesses a total amount normalized to one (the whole pie), and has the option of paying a recipient an arbitrary fraction $0 \leq \pi \leq 1$ of this pie. Contingent on this payment, the recipient obtains a payoff amounting to $p\pi$, where $p > 0$ is the 'multiplier of redistribution' (the inverse $1/p$ is also referred to as the 'price of redistribution'). If the dictator has CES preferences, given by Equation (1), where $o = p\pi$ and $s = (1 - \pi)$, his utility function can be written as

$$u(\pi) := \left[c(p\pi)^{\rho} + (1 - c)(1 - \pi)^{\rho}\right]^{\frac{1}{\rho}}. \tag{2}$$

The dictator determines his optimal giving by maximizing this one-dimensional function $u(\pi)$.

2.3. Interactive dictator games

In order to introduce the interactive setting, we assume that N players $(1, 2, \ldots, N)$ are arranged in a 'loop' of size N, see Fig. 1c. That is, Player 1 is the dictator of Player 2, who is the dictator of Player 3, ..., who is the dictator of Player N, who is the dictator of Player 1.

Each player initially possesses a total amount of one and may choose –simultaneously to all others– to make a payment of $0 \leq \pi_i \leq 1$ to his/her recipient $(i = 1, \ldots, N)$. As in the non-interactive case, this amount is multiplied by $p > 0$ so that Player i obtains an amount $p\pi_{i-1}$ from his/her predecessor, Player $i-1$, in the loop (the predecessor of Player 1 is Player N). Hence, Player i's total payoff consists of what he/she keeps and what he/she receives, amounting to

$$s_i := 1 - \pi_i + p\pi_{i-1}.$$

Next, we want to model how Player i cares for the final payoffs of other players. A canonical assumption, and the one we make here, is that Player i –rather than caring about every single other player separately– considers some form of aggregation of others' payoffs as given by a distribution of 'weightings' $\{w_j^{(i)}\}_{j \neq i}$ such that $w_j^{(i)} \geq 0$ for all $j \neq i$ and $\sum_{j \neq i} w_j^{(i)} = 1$,

$$\begin{aligned} o_i &:= \sum_{j \neq i} w_j^{(i)} s_j \\ &= \sum_{j \neq i} w_j^{(i)} (1 - \pi_j + p\pi_{j-1}). \end{aligned}$$

Different ways to distribute weightings $\{w_j^{(i)}\}_{j \neq i}$ express various plausible 'bracketing' assumptions regarding Player i's way of thinking about the other players; i.e. who is relevant for his/her decision.[11] For example, one plausible bracketing assumption is $w_j^{(i)} = \frac{1}{N-1}$ for all $j \neq i$, which corresponds to a 'broad' bracketing in the sense that all other players' payoffs matter equally to Player i. Also plausible would be a 'narrow' bracketing with $w_{i+1}^{(i)} = 1$ and $w_j^{(i)} = 0$ for $j \neq i + 1$, where Player i considers only his/her immediate recipient (i.e. the direct successor in the loop).[12]

It then follows that, given a weighting $\{w_j^{(i)}\}_{j \neq i}$, Player i's utility in the interactive setting reads as, cf. Equation (1),

$$\begin{aligned} u_i(\pi_1, \ldots, \pi_N) &= \left[c_i o_i^{\rho_i} + (1 - c_i) s_i^{\rho_i}\right]^{\frac{1}{\rho_i}} \\ &= \left[c_i \left(\sum_{j \neq i} w_j^{(i)} (1 - \pi_j + p\pi_{j-1}) \right)^{\rho_i} + (1 - c_i)(1 - \pi_i + p\pi_{i-1})^{\rho_i} \right]^{\frac{1}{\rho_i}}. \end{aligned} \tag{3}$$

[11] We are grateful to the associate editor and a reviewer for suggesting this model and terminology.

[12] We thank the associate editor for the observant remark that related bracketing assumptions could actually even matter in non-interactive dictator game implementations. Indeed, instead of caring exclusively about one's 'own' recipient, a dictator in a non-interactive setting might care about the distribution of payoffs in the game more generally, e.g. by considering the average recipient payoff. The giving decisions of all dictators would then constitute a single proper game. Future work should consider this possibility too.

Importantly, as opposed to the non-interactive case, this expression depends not only on Player i's own payment decision but also on those of other players. We stress that all player make their payment decision at the same time and thus without knowing the decisions of others. It follows that we are faced with a proper, i.e. interactive, one-shot/simultaneous move game and hence shall be interested in computing its Nash equilibria.

3. Results

In this section, we develop the rational-choice predictions for both types of settings. We first identify the optimal payments for non-interactive dictator games. Second, we contrast these predictions with pure-strategy Nash equilibrium benchmarks under complete information in the interactive setting.

3.1. Non-interactive dictator games

Starting from Equation (2), the utility-maximizing payment π^* is readily computed (see also Jakiela 2013) from the first-order condition $u'(\pi^*) = 0$ and yields, for $\rho < 1$,

$$\pi^* = \frac{1}{1 + p\left(\frac{1-c}{pc}\right)^{\frac{1}{1-\rho}}} =: \frac{1}{1+pa}. \tag{4}$$

Here, a has the meaning of an efficiency-adjusted selfishness parameter: it is decreasing in c and increasing (decreasing) in ρ for $p < \frac{1-c}{c}$ ($p > \frac{1-c}{c}$). As a consequence, we obtain the intuitive interpretation that, first, more concern (higher c) yields higher payments, and, second, more efficiency orientation (higher ρ) yields lower payments for low redistribution multipliers and higher payments for high redistribution multipliers. Equation (4) also shows that the optimal payment is increasing in p for $\rho > 0$, decreasing in p for $\rho < 0$, and independent of p for $\rho = 0$.[13]

The limiting cases of Formula (4) relate to those of $v(o, s)$ that were discussed above as follows.

Perfect egoist. If $c = 0$, optimal payment is $\pi^* = 0$ for all ρ.
Perfect altruist. If $c = 1$, optimal payment is $\pi^* = 1$ for all ρ.
Rawlsian. If $\rho \to -\infty$ and $c \neq 0, 1$, then $\pi^* = \frac{1}{1+p}$, and $1 - \pi^* = p\pi^*$: equality adjusted for p.
Efficiency optimizer. If $\rho \nearrow 1$ and $c \neq 0, 1$, then $\pi^* = 0$ ($\pi^* = 1$) if $p < \frac{1-c}{c}$ ($p > \frac{1-c}{c}$).

3.2. Interactive dictator games

Best replies

In order to determine the Nash equilibria in pure strategies of the interactive dictator game, we first compute the best reply $\mathcal{BR}_i(\pi_{-i})$, where π_{-i} is shorthand notation for $(\pi_1, \ldots, \pi_{i-1}, \pi_{i+1}, \ldots, \pi_N)$ for each Player i starting out from his/her utility function $u_i(\pi_1, \ldots, \pi_N)$, see Equation (3). We present here the analysis of the cases excluding limiting preferences discussed above (perfect egoist/altruist, Rawlsian, and efficiency optimizer), which are treated in Appendix A.1. We thus assume

$$0 < c_i < 1 \text{ and } -\infty \neq \rho_i < 1. \tag{5}$$

For the first order condition for a best reply, it suffices to consider $u_i(\pi_1, \ldots, \pi_N)^{\rho_i}$, even for $\rho_i < 0$ (the case $u_i(\pi_1, \ldots, \pi_N) = 0$ corresponds to a minimum and can thus be excluded). We obtain

$$\begin{aligned} & 0 = \partial_{\pi_i}(u_i^{\rho_i}) = \partial_{\pi_i}\left[c_i o_i^{\rho_i} + (1-c_i)s_i^{\rho_i}\right], \\ \Leftrightarrow\ & w_{i+1}^{(i)} p c_i o_i^{\rho_i - 1} = (1-c_i)s_i^{\rho_i - 1} \\ \Leftrightarrow\ & s_i = \left(\frac{w_{i+1}^{(i)} p c_i}{1-c_i}\right)^{\frac{1}{\rho_i - 1}} o_i \\ \Leftrightarrow\ & s_i =: a_i o_i, \end{aligned} \tag{6}$$

where all functions are evaluated at $(\pi_1, \ldots, \pi_{i-1}, \mathcal{BR}_i(\pi_{-i}), \pi_{i+1}, \ldots, \pi_N)$. It is readily verified that the Cobb-Douglas case, $\rho_i = 0$, leads to the same result. Relation (6) has the intuitive interpretation that Player i's best reply yields higher (lower) payoffs for himself than for the weighted average of the others if $a_i > 1$ ($a_i < 1$), where a_i is the efficiency-adjusted

[13] The latter observation is consistent with the fact that in a Cobb-Douglas utility function, p only enters as an overall multiplicative constant. It therefore has no influence on optimal payment, and the utility-maximizing payment is $\pi^* = c$, independent of the multiplier of redistribution $p > 0$.

selfishness parameter in the interactive case, see Equation (4).[14] In particular, there are no dominant strategies as long as $0 < a_i < \infty$, because the giving constituting a best reply will generically depend on others' giving.

Nash equilibria: general results

We use the best reply correspondences to compute the Nash equilibria of the N-player interactive dictator game. The next theorem is our main result for this case. It establishes that, for a wide range of other-regarding preferences, extremal payments constitute Nash equilibria in the interactive setting.

Theorem 1. *The N-player interactive dictator game admits the following equilibria $(\pi_1^*, \dots, \pi_N^*)$ for arbitrary bracketing weights $\{\{w_j^{(i)}\}_{j\neq i}\}_{i=1}^N$ with $\sum_{j\neq i} w_j^{(i)} = 1$ for all i:*

1. *If for all i either $0 \leq a_i \leq 1$, or $\rho_i = 1$ and $w_{i+1}^{(i)} p c_i = (1 - c_i)$, then $(\pi_1^*, \dots, \pi_N^*) = (1, \dots, 1)$ is an equilibrium.*
2. *If for all i either $1 \leq a_i \leq \infty$, or $\rho_i = 1$ and $w_{i+1}^{(i)} p c_i = (1 - c_i)$, then $(\pi_1^*, \dots, \pi_N^*) = (0, \dots, 0)$ is an equilibrium.*

This result stands in stark contrast to the non-interactive setting where intermediate payments are made whenever $0 < a < \infty$. The extremal giving structure of Theorem 1 is quite robust with respect to variations of the underlying modeling assumptions. For one, zero payments also emerge in equilibrium with respect to other well-known specifications of other-regarding preferences (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002). Moreover, the same kind of equilibrium structure persists when information regarding other players is modeled as incomplete: for instance, provided that beliefs (including higher-orders) attach non-zero probability only to $a_i \geq 1$, equilibrium payments remain zero. For further details, generalizations, as well as for a proof of Theorem 1, the reader is referred to the Appendix A.

To help gain some intuition for our result, consider Equation (6): if all players are, for instance, sufficiently selfish, i.e. if $a_i > 1$ for all i, then any player's best reply will ensure that others have less final payoff than him/her. Iteratively eliminating strictly dominated strategies then leads to the only fixed point in best replies, which is the Nash equilibrium with zero-giving for all players.[15] Similarly, if instead all players are somewhat altruistic, i.e. if $a_i < 1$ for all i, then an analogous argumentation leads to full-giving by all players in equilibrium.

To express our findings in terms of our initial parameters c_i, ρ_i and $w_{i+1}^{(i)}$, we point out that the condition $a_i \gtrless 1$ is equivalent to

$$c_i \gtrless \frac{1}{1 + w_{i+1}^{(i)} p} =: c_{\text{crit}}. \tag{7}$$

Note that the condition for zero payments, $c_i < c_{\text{crit}}$, is easier to fulfill the lower the multiplier of redistribution p and the smaller the weight $w_{i+1}^{(i)}$. An analogous statement can be made if $c_i > c_{\text{crit}}$. From a welfare perspective, this p-dependence can thus be seen as an (imperfect) tendency towards efficiency in interactive designs in comparison with non-interactive ones. Note however that Condition (7) does not depend at all on the players' respective efficiency orientations in our CES-based model.[16] This, again, is opposed to the non-interactive setting.

While the existence of zero-giving equilibria for not purely selfish players in itself already clashes with findings in the traditional non-interactive approach, uniqueness of the equilibria described in Theorem 1 can in addition be proved depending on the underlying bracketing assumptions. We explicitly analyze the following canonical cases.

Narrow bracketing. In this case $w_{i+1}^{(i)} = 1$ and $w_j^{(i)} = 0$ for all i and $j \neq i + 1$ and Condition (7) simplifies to

$$c_i \gtrless \frac{1}{1 + p} =: c_{\text{crit}}^{\text{narrow}}, \tag{8}$$

which corresponds to the critical value for monotonicity in ρ in the non-interactive case, cf. Equation (4). The case of narrow bracketing also allows us to think about welfare comparison with the non-interactive setting since there the bracketing trivially concerns only one subject, i.e. the recipient (however beware the caveat mentioned in Footnote 12). If, for instance, we assume $c_i < c_{\text{crit}}$, we may conclude that an interactive implementation is more efficient for $p < 1$ while a non-interactive implementation is for $p > 1$. An analogous statement can be made if all $c_i > c_{\text{crit}}$.

Broad bracketing. In this case $w_j^{(i)} = 1/(N - 1)$ for all i and $j \neq i$, so that we obtain the modified bounds

$$c_i \gtrless \frac{1}{1 + \frac{p}{N-1}} =: c_{\text{crit}}^{\text{broad}}. \tag{9}$$

[14] We show in Appendix A.1 that most combinations of the excluded parameter choices, specifically $c_i = 0, 1$ and $\rho_i = -\infty, 0, 1$, can be absorbed by values of $a_i \in [0, \infty]$, so that Equation (6) remains valid. Only the case where $\rho_i = 1$ and $w_{i+1}^{(i)} p c_i = (1 - c_i)$ requires a separate treatment.

[15] This is equally true for a finite (as encountered in the economic laboratory) and a continuous (as in our theory formulation) strategy space (Chen et al., 2007).

[16] This is even true for the special case $\rho_i = 1$ since $c_i > c_{\text{crit}}$ ($c_i < c_{\text{crit}}$) is then absorbed in $a_i = 0$ ($a_i = \infty$).

Thus, $c_i < c_{\text{crit}}^{\text{broad}}$ is then satisfied even for substantially altruistic players if p/N is sufficiently small. In particular, we obtain a zero-giving equilibrium under broad bracketing as long as the players are not perfectly altruistic and sufficiently numerous.

Under these bracketing assumptions we obtain the following uniqueness result.

Theorem 2. *The following statements hold for the N-player interactive dictator game under both narrow and broad bracketing:*

1. *If* $0 \leq a_i < 1$ *for all i, then* $(\pi_1^*, \ldots, \pi_N^*) = (1, \ldots, 1)$ *is the unique pure-strategy equilibrium.*
2. *If* $1 < a_i \leq \infty$ *for all i, then* $(\pi_1^*, \ldots, \pi_N^*) = (0, \ldots, 0)$ *is the unique pure-strategy equilibrium.*

The proof is given in Appendix A.5.

Nash equilibria: 2-player case

We conclude this section with a detailed discussion of the case $N = 2$, where a complete description of all pure strategy Nash equilibria turns out to be technically feasible. Note that in this case all possible bracketing assumptions coincide.

Proposition 3. *For* $a_i \in [0, \infty]$, *the pure strategy equilibria* (π_i^*, π_{-i}^*) *of the 2-player interactive dictator game can be characterized as follows:*

1. *If* $a_i a_{-i} \neq 1$ *(generic case) the equilibrium strategy combination is unique and located on the boundary of the unit square* $[0, 1]^2$. *More precisely, we have:*
   i. *If* $a_i a_{-i} > 1$ *and* $a_{-i}, a_i \geq 1$ *then* $(\pi_i^*, \pi_{-i}^*) = (0, 0)$.
   ii. *If* $a_i a_{-i} > 1$ *and* $a_{-i} > 1 > a_i$ *then* $(\pi_i^*, \pi_{-i}^*) = (\frac{1-a_i}{1+pa_i}, 0)$.
   iii. *If* $a_i a_{-i} < 1$ *and* $a_{-i}, a_i \leq 1$ *then* $(\pi_i^*, \pi_{-i}^*) = (1, 1)$.
   iv. *If* $a_i a_{-i} < 1$ *and* $a_{-i} < 1 < a_i$ *then* $(\pi_i^*, \pi_{-i}^*) = (\frac{1+p}{1+pa_i}, 1)$.
2. *If* $a_i a_{-i} = 1$ *the equilibrium strategy combinations are given by a straight line segment. More precisely, for* $a_{-i} \geq 1 \geq a_i$, *we have*

$$\left\{ (\pi_i, \pi_{-i}^*) = \left(\frac{1 - a_i + (a_i + p)t}{1 + pa_i}, t \right) \middle| 0 \leq t \leq \frac{a_i + pa_i}{a_i + p} = \frac{1 + p}{1 + pa_{-i}} \right\}.$$

3. *If* $a_i = 0$ *and* $a_{-i} = \infty$, *the unique equilibrium strategy combination is* $(\pi_i^*, \pi_{-i}^*) = (1, 0)$, *consistent with 1.ii and 2.*
4. *If* $\rho_i = 1$ *and* $c_i p = 1 - c_i$, *then any* (π_i^*, π_{-i}^*) *with* $\pi_{-i}^* = \mathcal{BR}_{-i}(\pi_i^*)$ *is an equilibrium strategy combination. In particular, if also* $\rho_{-i} = 1$ *and* $c_{-i} p = 1 - c_{-i}$, *then any* (π_i^*, π_{-i}^*) *is an equilibrium strategy combination.*

The proof is given in Appendix A.6. While Theorem 1 and 2 are concerned with the generic extremal payments we highlight two special cases of intermediate payments treated in Proposition 3 which are readily generalized to the N-player case. First, $a_i a_{-i} = 1$ may be observed experimentally, for instance in a basic dictator game ($p = 1$) where both players value their own and other's payoffs equally ($c_i = c_{-i} = \frac{1}{2}$, $\rho_i, \rho_{-i} < 1$); in which case $a_i = a_{-i} = 1$. In a non-interactive setting, such a dictator would split 50-50. In an interactive setting, however, any split of the pie, as long as both players split it the same way, constitutes a Nash equilibrium, i.e. $(\pi_i^*, \pi_{-i}^*) = (t, t)$, $0 \leq t \leq 1$, which is consistent with Item 2 above. Using the best reply correspondence, Equation (10) in the Appendix, this equal-payment equilibrium, follows immediately also for an arbitrary number of players with such preferences given $p = 1$.

Second, if in a basic dictator game ($p = 1$) a player is perfectly efficiency-oriented ($\rho_i = 1$), and has as much concern for himself/herself as for his opponent ($c_i = c_{-i} = \frac{1}{2}$), then his/her preferences coincide with those of a social planner who aims to maximize total wealth. Thus, the player is indifferent with respect to his/her payment; just as described by Item 4 above. This is equally true in the N-player setting as long as concerns are modified to account for bracketing weights, $c_i = 1/(1 + w_{i+1}^{(i)})$, see the efficiency optimizer case discussed in Appendix A.1.

4. Experimental evidence

4.1. Existing studies

Dictator games are one of the fruit flies of experimental economics. Despite fundamental differences that result in terms of strategic incentives non-interactive and interactive protocols have been used interchangeably. The pioneering experiments by the 'inventors', Kahneman et al. (1986) (without monetary incentives) and Forsythe et al. (1994) (with monetary incentives), were completely non-interactive. These studies featured relatively small subject samples, and their main purpose was to test the hypothesis that all subjects were narrowly motivated by material self-interest, falsified by the fact that substantial percentages of the samples gave positive amounts.

However, allocating half of the subjects (the recipients) to a purely passive role is extremely unattractive from a cost perspective, because those subjects take absolutely no decision and, thus, do not generate any exploitable experimental data.[17] Perhaps because of that, many of the more recent experiments –which recruited larger number of subjects, because they were designed to go beyond constituting counterexamples against the narrow self-interest hypothesis– opted for interactive experimental protocols. This leads to all subjects generating data as everybody then acts as both dictator and recipient at the same time, but also comes at the hitherto unacknowledged 'cost' of introducing unwanted strategic incentive. Typically, and presumably with the intent of making the link between giving and taking less immediately evident, any two subjects are precluded from being mutual dictators by virtue of the matching protocol. Instead, 'loops' with sizes larger than three are being implemented to incentivize decisions (see Fig. 1c).[18]

As mentioned in the Introduction, existing dictator game studies have to the best of our knowledge not thoroughly distinguished between the two implementations, despite the fundamental differences in terms of incentives. No study following an interactive experimental design has considered decision inter-dependencies –neither theoretically by developing the rational-choice benchmarks (i.e. Nash equilibrium) nor experimentally by controlling satisfactorily for strategic sophistication. What is more, we know of only one single study that compared, at least experimentally, the two protocols using a randomized between-subject design (Korenok et al., 2013).[19] This study, in a relatively small subject sample ($N = 57$ (non-interactive) resp. $N = 62$ (interactive) vs. $N = 206$ (both treatments) in our experiment), results in no conclusive evidence. Furthermore, its main focus is on endowment effects –a parameter which we held fixed in our own experiment to avoid further confounds– and an appropriate control for beliefs has been left aside. To be clear, theirs is a perfectly legitimate finding, but as we lay out in Appendix B even in this case an interpretational ambiguity persists in view of our theoretical findings.

A particularly instructive case that illustrates how our theoretical findings may be relevant is that of Fisman et al. (2015b) where coincidentally both interactive and non-interactive protocols were used. In that study, it is argued that Yale 'elite' students (where an interactive protocol was used) are more selfish and efficiency-oriented than 'average' citizens in the US (where a non-interactive protocol was used).[20] These findings are then suggested as an explanation for the growing inequality in the US wealth distribution as being potentially driven by policies made by that very elite. This may well be true, but, as we have argued above, cannot be concluded from their data prima facie, as the game structure of the experiment remains unacknowledged: the role of strategic incentives, awareness, and sophistication is not controlled for, and this means that the more extremal decisions that characterize the Yale 'elite' sample compared with the 'average' sample may be a result of strategic motives as well.[21] Fig. 3 illustrates the differing giving behavior of the two subject pools. In order to visualize decisions on the level of subjects, we create for each Subject i an extremality index $\phi_i \in [0, 1]$ (horizontal axis) based on the absolute distance from equal-splitting over all decisions, $\phi_i := \frac{1}{M}\sum_m 2\left|\pi_i^{(m)} - \frac{1}{2}\right|$, where $m \in \{1, \ldots, M\}$ labels the decisions. Specifically, the two extremes $\phi_i = 0$ and 1 respectively represent equal-splitting and extremal-giving, that is, zero- or full-giving in every single decision by a given Subject i. For a graphical representation on the level of individual decisions, see Fig. 5 in Appendix D. We caution the reader that a direct comparison with the specific equilibrium structure as in Theorem 1 requires some care. In order to test for GARP violations, Fisman et al. (2015b) are by design forced to vary endowments – an additional parameter that our model and experiment keep fixed so as to exclude endowment effects as an explanatory variable. Moreover, their experimental design involves considerable uncertainty. While non-trivial zero-payment equilibria still exist in their setting, they are perhaps less generic (cf. the generalization of Theorem 1 to settings with uncertainty in Appendix A.3).[22] Notwithstanding, this still shows that optimal play in interactive settings is very different from non-interactive settings.

[17] The cost issue is somewhat mitigated when experiments are conducted online, and there are some recent studies conducting non-interactive online experiments with samples representing the general population, see e.g. Fisman et al. (2017).

[18] The two main variants of the matching protocol are as follows. Either a randomly ordered loop involving all session subjects is generated (see e.g. Murphy et al. 2011), or, following Andreoni and Miller (2002), participants are randomly matched under the sole restriction that no two subjects can be each others' dictators (or recipients, respectively). Depending on session size, the latter typically leads to several disjoint dictator loops of varying sizes that are smaller than the whole session population. Loop size has no impact on Nash equilibrium predictions in Theorem 1 given narrow bracketing, but it does given broad bracketing. Proposition 4 in Appendix A.7 shows that, given typical experimental parameters, it is likely for a given player to be in a sufficiently long loop, so that zero-giving remains the (unique) Nash equilibrium.

[19] Greiff et al. (2018) measure within-subject differences in Social Value Orientation (SVO) when interactive implementations are made explicit using a matrix formulation. A within-subject design is, however, unsuitable for our purposes since the actual game being played there is a combination of the non-interactive and interactive games. This results in a third, combined, game that is based on yet another set of strategic incentives, so that giving in the non-interactive (interactive) stage is not equivalent to giving in a stand-alone non-interactive (interactive) game.

[20] The paper makes comparisons by (re-)analyzing data from an interactive Yale Law School study (Fisman et al., 2009), from an interactive Berkeley study (Fisman et al., 2007), and from a non-interactive American Life Panel study (Fisman et al., 2017).

[21] To be clear, we make no claim here regarding whether or not elites are more selfish than the general population. In a related study (Li et al., 2017), where both future physicians (another 'elite') and average citizens were offered to give in a non-interactive setting, it was found that this elite sample made more selfish and efficiency-oriented decisions than the general population sample.

[22] Interestingly, their data shows more extremal payments under the interactive protocol than under the non-interactive one – as would be in alignment with Theorem 1. Whether Nash equilibrium drives these decisions is of course speculative at this point and an 'inverted' control experiment with subjects from the same two subject pools would be required for a proper test.

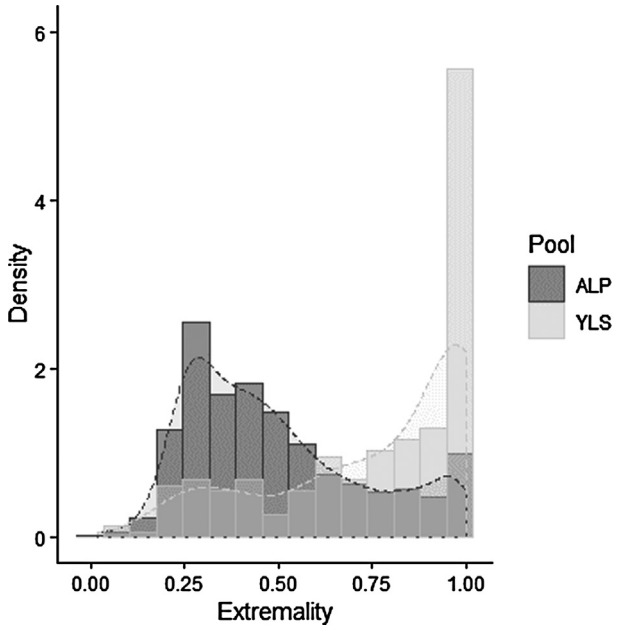

Fig. 3. Extremality of decisions of 'elite' versus 'average' in Fisman et al. 2015b. *Non-interactive (ALP):* Subjects are representative of the US public and were recruited via the American Life Panel (ALP). Giving is mostly intermediate with almost no equal-splitting and few extremal payments. *Interactive (YLS):* Subjects were students at Yale Law School (YLS). The giving pattern is extremal (and quite unusual compared with previous studies) with fewer intermediate payments and almost no equal-splitting. *Source:* authors' own figures based on data obtained from Fisman et al. (2015b) and from the ALP, as per Science's data policy.

4.2. Our experiment

Motivated by our discovery that interactive and non-interactive protocols are being used without distinction in the experimental literature (see the extended abstract (Grech and Nax, 2017)), we also set out to understand how these protocols matter behaviorally. We conducted a tailor-made experimental test akin to a randomized controlled trial (RCT).[23] In order to avoid any unwanted 'researcher degrees of freedom' such as the possibility of ad hoc and ex post data rationalizations (Simmons et al., 2011), our analysis plan (including a complete description of the hypotheses being tested and of the statistical procedures being used) was pre-registered with the Open Science Framework (OSF) prior to data collection. This commits us to perform and report on all pre-registered analyses in our paper, while labeling any additional analyses explicitly as exploratory.

Hypotheses

We shall focus on the following conceptual hypotheses regarding non-interactive vs. interactive dictator games.

(H0) There are no protocol differences.
(H1) Protocol differences exist
  (H1.h0) …and can be explained by rational-choice benchmarks.[24]
  (H1.h1) …and *cannot* be explained by rational-choice benchmarks.

We point out that testing *(H0)* against *(H1)*, independently of the secondary hypotheses, tests directly for the experimental non-interchangeability of non-interactive and interactive protocols – independent of our theoretical results. Conversely, it is straightforward to see that regardless of the experimental outcome, interchanging protocols is hard to justify (a detailed theoretical discussion of the hypotheses' implications is given in Appendix B). That is, the theoretical and experimental approach stand for two independent tests of the present paper's central premise.

Experimental design

We conducted a randomized between-subject experiment, where the only treatment variation was whether an interactive or a non-interactive protocol was used. Our instructions were adapted from Andreoni and Miller (2002) and were identical

[23] We thank the (advisory) editors and reviewers who motivated us to pursue this direction.
[24] Recall that *(H1.h0)* posits that non-interactive treatments induce predominantly intermediate payments, while interactive implementations are characterized by extremal payments.

Table 1
Summary Statistics. Mean and standard deviation of giving, and proportion of extremal giving by treatment, gender and self-reported bracketing ('narrow' means one other person is included in a subject's decision-making).

	All	Int.	Non-int.	Female	Narrow
Mean giving	289.3	313.3	265.4	302.1	313.3
Median giving	240	200	300	250	300
Standard deviation	269.5	268.5	268.3	270.0	259.8
Zero-giving (%)	17.8	13.2	22.5	14.9	14.8
Half-giving (%)	10.7	13.2	8.3	11.1	12.6
Full-giving (%)	4.2	4.5	3.9	4.7	3.9
Number of obs.	8'240	4'120	4'120	5'040	3'660

for both treatments (to avoid further framing effects) except for the information that was provided regarding which protocol (interactive or non-interactive) was being played and the fact that we kept budgets fixed in order to avoid confounds related to endowment effects.

As mentioned above, the experiment was fully pre-registered at the OSF before data collection. Subjects were recruited from our lab's (ETH Zurich's Decision Science Laboratory also known as DeSciL) pre-registered pool from among the general workforce on Amazon's Mechanical Turk (MTurk). Those subjects signed up to the pool and are aware of the scientific principles of the administering laboratory (including anonymity, no deception, etc.). We opted for an online experiment to minimize other treatment differences such as the presence of numerous passive subjects in a laboratory. Our online subjects explicitly consented to participate, and, as was pre-registered, no specific set of subjects was excluded except for those who took part in dictator game experiments with our laboratory during the three months preceding our experiment. The experiment was run as a Qualtrics study following the Operational Rules which commit the laboratory to modern experimental economics standards including no deception, anonymity, right of termination, etc., as confirmed by an Institutional Review Board Certificate of the German Association for Experimental Economic Research.[25]

Our experiment involved a total of 618 participating subjects. 412 subjects acted in the role of a dictator (206 per protocol). Another 206 subjects were randomly selected as recipients in the non-interactive protocol, directly taken to a short exit survey after reading the instructions, and paid without making any payout-relevant decisions.[26] The experiment was run on November 21, 2017, between 3pm and 7pm Eastern Standard Time. Subjects recruited during that period were randomly allocated to one of the three possible roles: the dual dictator-and-recipient role in the interactive protocol, the dictator role in the non-interactive protocol, or the recipient role in the non-interactive protocol.

Each dictator took 20 decisions. Each decision involved determining how many of 1'000 tokens to keep, and how many to give. The 20 decisions varied with respect to the multiplier of redistribution p, which was set to values 0.1, 0.2, 0.3,..., 2.0, presented to each dictator in random order.[27] Earnings included a fixed show-up fee of 1 USD irrespective of position, and a variable bonus that depended on the roles and decisions of the subjects. One decision was randomly selected per dictator and paid out: in the non-interactive protocol, a randomly selected recipient was paid and each recipient was matched with exactly one dictator; in the interactive protocol, each subject was the dictator of someone and the recipient of someone else. Each 200 tokens earned that way were worth 1 USD. Hence, possible bonus payments operated in the range of 0 USD to 15 USD, depending on role, treatment, and randomly selected payments. The experiment took between 1 minute and 30 minutes, depending on subject and role. Actual bonus payments varied between 0 USD and 12.50 USD, with an average earning of 3.27 USD, which is well above standard MTurk wages.

Results

Descriptive summary statistics. Table 1 gives a descriptive summary of the experimental data.[28] Overall, observed giving decisions in the interactive case were on average higher and resulted in less zero giving than in the non-interactive case.[29]

Hypotheses testing. In what follows, we provide inferential statistics for differences between the non-interactive and the interactive treatment. As committed in our pre-registration, we interpret findings to be significant as per the 95% confidence level. In addition, we refer to findings that are significant at the 90% but not at the 95% confidence level as 'marginal', and to findings that are not significant at the 90% confidence interval as 'insignificant'.

Hypothesis (H0) is rejected: We observe significantly different giving distributions overall when comparing the two treatments (Kolmogorov-Smirnov test: p-value < 0.001, Mann-Whitney-Wilcoxon test: p-value < 0.001), as shown in Table 3 in

[25] The Expedite Institutional Review Board Certificate Number 1 (Code RJPwvcoT), the experimental instructions and all other materials are deposited at osf.io/fsg52.

[26] In addition, 14 subjects who were recruited as passive recipients were paid the average dictator decision, because their dictators were lost due to internet disconnections before any decision was taken. Hence, one session included only 6, not 20 active dictators.

[27] This represents a deviation from the Andreoni and Miller (2002)-protocol which considers intersecting budget lines and is therefore forced to vary endowments. In order to be sure that potential differences would not be driven by endowment effects, we kept endowments fixed.

[28] See also Figs. 6 and 7 in Appendix D for visualizations.

[29] Further summary related to giving decisions include an average time of decision of 9.1 seconds (s.d. 16.2), and skewness and kurtosis of 0.926 and 3.274, respectively (0.825 and 3.138 in non-interactive versus 1.048 and 3.499 in interactive).

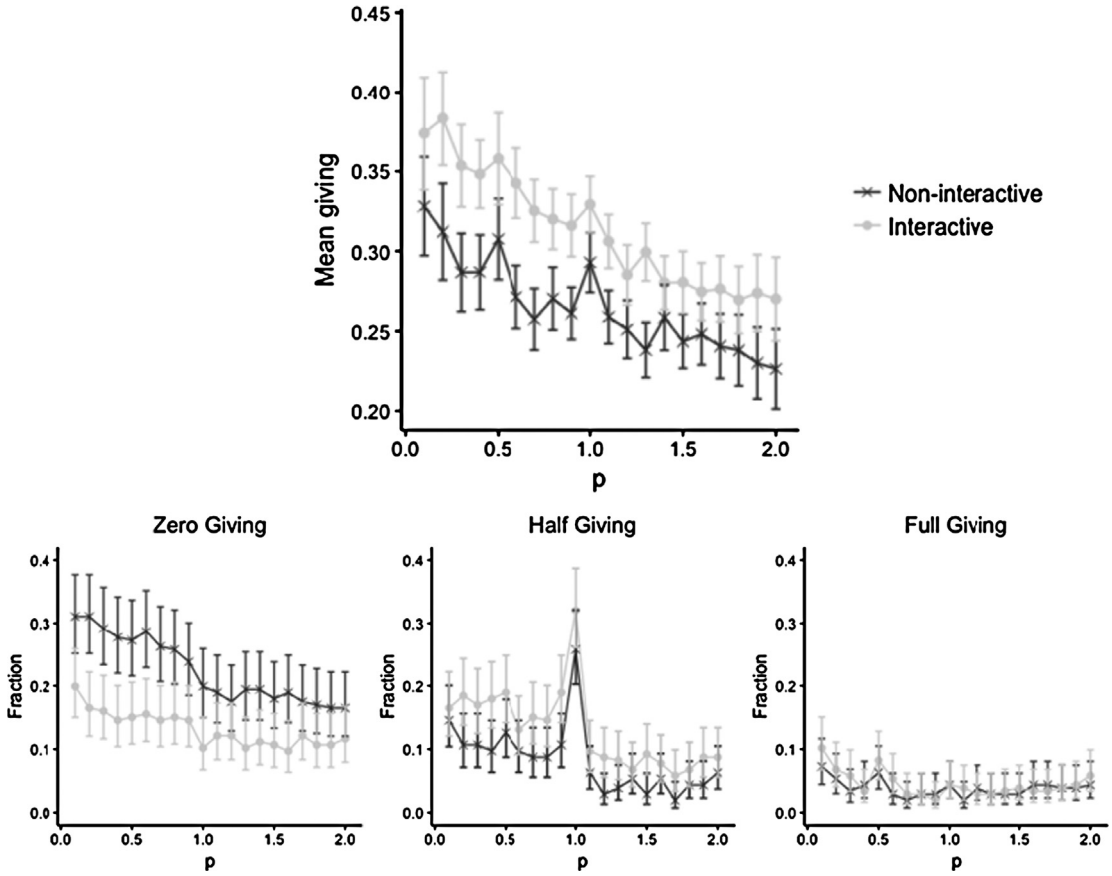

Fig. 4. Top: Overall Giving. Average giving, in the range $\pi \in [0.2, 0.45]$, decreases in the multiplier of redistribution p. Error bars denote the 95% confidence interval corrected for between-subject differences (Cousineau et al., 2005; Morey et al., 2008). Bottom: Extremal giving. Fraction of extremal giving displayed in the range $[0.2, 0.45]$, in dependence of $p \in [0, 2]$. Error bars denote the 95% confidence interval based on the Wilson score-interval for binomial proportions. *Zero Giving:* Zero giving decreases with growing p. Overall, there is more zero giving in the non-interactive protocol. *Half giving:* Half giving peaks at the standard dictator game $p = 1$. *Full giving:* The fraction of full giving is comparably low in both treatments.

Appendix C. With respect to variance levels, there are no significant protocol differences (Levene's test: p-value $= 0.928$, Brown-Forsythe tests: p-value $= 0.965$). Giving is significantly higher under the interactive than under the non-interactive protocol, which can also be inferred from an estimated treatment effect of over twenty percent relative to the intercept in our exploratory OLS – see Table 4 in Appendix C, which also provides additional details concerning giving patterns with respect to further controls. Under both protocols, giving decreases in the multiplier of redistribution, as is illustrated in Fig. 4 (upper panel) and our exploratory OLS. With regards to distributional treatment differences dependent on individual multipliers of redistribution p, Table 3 indicates that treatment effects are predominantly significant for multipliers of redistribution below 1.0 (only exception: $p = 0.1$). For 1.0 and above significant differences occur less frequently (no significance for $p = 1.0, 1.2, 1.4, 1.6, 1.8$). We detect no significant differences in variance levels between the protocols for any level of p.

Hypothesis (H1.h0) is rejected: Nash equilibrium is characterized by extremal decisions, which is not confirmed by the overall shape of the data. This reflects the well-established fact that uninformed participants oftentimes deviate from Nash equilibrium behavior in one-shot games (see also Footnote 8). In order to test explicitly whether extremal decisions are more frequent in the interactive protocol, we perform probit regressions, see Table 2 which also includes information on the focal point of half giving. We find significant treatment differences regarding zero-giving (more zero-giving in non-interactive). No significant differences are found regarding full-giving. Note also that half-giving is marginally significantly more frequent in the interactive protocol.

The dependency of extremal and focal giving on the multiplier of redistribution p can be summarized as follows: zero- and half-giving decrease significantly in p in both treatments – however, with half-giving spiking notably at $p = 1$. There is no significant correlation between full-giving and p. Fig. 4, lower panel, illustrates these trends.

Robustness: The above effects relating to treatment differences regarding extremal decisions are robust with respect to the inclusion of additional controls. The additional effects summarize as follows. Effects regarding subjects' age and order of decision-making are insignificant. Subjects' gender is significant inasmuch as that women are less likely to give zero. There

Table 2
Extremal and Half-giving Decisions. Estimates of the ordered probit regressions with individual-level clustering. Parentheses indicate [robust standard errors] and (p-values).

	(1) Zero-giving		(2) Half-giving		(3) Full-giving	
	(Reduced)	(Controls)	(Reduced)	(Controls)	(Reduced)	(Controls)
Interactive	−0.465*	−0.424*	0.256	0.250	0.159	0.070
	[0.141]	[0.148]	[0.189]	[0.134]	[0.132]	[0.192]
	(0.001)	(0.004)	(0.052)	(0.061)	(0.401)	(0.715)
Inter.$\times p$	−0.018*	−0.019*	−0.034*	−0.033*	−0.017	−0.018
	[0.007]	[0.007]	[0.010]	[0.007]	[0.007]	[0.010]
	(0.005)	(0.004)	(<0.001)	(<0.001)	(0.112)	(0.092)
Non-int.$\times p$	−0.028*	−0.031*	−0.035*	−0.035*	−0.008	−0.008
	[0.005]	[0.006]	[0.011]	[0.007]	[0.007]	[0.011]
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.476)	(0.491)
Order		0.002		−0.002		0.003
		[0.002]		[0.002]		[0.003]
		(0.323)		(0.316)		(0.230)
Female		−0.303*		0.040		0.165
		[0.131]		[0.112]		[0.153]
		(0.021)		(0.721)		(0.281)
Age		−0.005		0.001		−0.002
		[0.006]		[0.005]		[0.008]
		(0.442)		(0.783)		(0.857)
Self-centrism		0.879*		−0.283		0.101
		[0.171]		[0.159]		[0.252]
		(<0.001)		(0.075)		(0.689)
Narrow bracketing		−0.045		0.086		0.008
		[0.151]		[0.122]		[0.175]
		(0.767)		(0.481)		(0.964)
Interactive play		0.066		−0.045		0.394*
		[0.174]		[0.123]		[0.192]
		(0.705)		(0.718)		(0.040)
(Cut)	0.467	0.301	1.044	1.099	1.689	1.848
	[0.093]	[0.278]	[0.146]	[0.216]	[0.097]	[0.339]
Log pseudolikelihood	−3763.3	−3513.3	−2725.6	−2712.4	−1423.5	−1397.8
Number of obs.	8'240	8'240	8'240	8'240	8'240	8'240
Number of ind.	412	412	412	412	412	412

* Stands for $p < 0.05$.

are no other significant gender differences in terms of overall giving, full-giving, and half-giving. Belief-related controls were generated from subjects' responses to questions we asked at the end of the experiment e.g. regarding how many other subjects they included in their considerations when making decisions. Self-centrism indicates that a subject considered only himself/herself. Narrow bracketing indicates that at most two other subjects were included – one recipient and one dictator. We find that whether subjects bracketed the decision narrowly or broadly has no significant effects, but subjects who do not include anyone else in their decision ('self-centric' bracketing) give less overall and zero more often (non-interactive versus interactive, respectively, we have 44.2 versus 44.7 percent narrow bracketing, and 18.9 versus 11.7 percent self-centrism.) Finally, certain subjects stated that they operated under an interactive frame of mind ('interactive play'), that is, they included others in their decision-making both in terms of that they believed to be affected by others and that others' decisions affected them (non-interactive versus interactive, respectively, we have 9.2 percent –incorrectly– versus 30.6 percent –correctly– operating under 'interactive play'). Those subjects are more likely to give everything, which is the one indicator that is line with Nash equilibrium benchmarks, which partially explains the overall higher giving in the interactive treatment. All other effects are insignificant. Assessing the robustness of our significant controls, we performed additional logit regressions including each control individually that had a significant effect in the probit regression including all controls. The effects remain significant. The reader is referred to online material for details (cf. Footnote 25) (see also (Mäs and Nax, 2016) for similar regressions).

Differences with Fisman et al. (2015b). A key difference is that in neither of our treatments we observe nearly as many extremal decisions as in the YLS data of Fisman et al. (2015b), cf. Fig. 3 and Fig. 6. Beyond population differences between our two studies, other differentiating factors include the experimental design and framing. Fisman et al. (2015b) consider randomly generated multipliers of redistribution up to $p = 10$, while our maximal value is $p = 2$. Hence, the relatively low rate of full-giving in our experiment may be due to the fact that full-giving becomes an evidently efficient thing to do

only when subjects face higher values of p. Another striking difference between the two datasets concerns the lack of a peak at the Rawlsian effective equal-splitting $\pi = 1/(1+p)$ in our data, which shows a peak at equal-splitting $\pi = 1/2$ instead (compare Figs. 5 and 7 in Appendix D). This is likely driven by another difference in experimental design; while we –following Andreoni and Miller (2002)– explain the multiplier of redistribution p in words, Fisman et al. (2015b) graphically illustrate effective earnings to participants.

Discussion

As detailed when stating our hypotheses, our experimental test questions existing practices and (implicit) assumptions in dictator game research by construction –regardless of the outcome– due to our theoretical results. Nevertheless, we can learn more from our specific experimental data under additional assumptions.

The rejection of Hypothesis *(H0)* results from observing that giving is significantly higher under the interactive than under the non-interactive protocol. Thus, if we *assume* humans to have absolutely no strategic awareness when playing interactive dictator games (despite all objections given earlier), then our experimental data suggests that previous studies relying on interactive protocols have overestimated pro-sociality. In addition, observing no significant differences regarding the frequency of full giving between protocols and even significantly less zero giving under the interactive protocol leads to the rejection of Hypothesis *(H1.h0)*. This implies that rational-choice benchmarks (optimal play of interacting and fully strategic utility-maximizers with other-regarding preferences) are not an appropriate model to describe behavior in experimental interactive dictator games. Since *(H0, H1.h0, H1.h1)* are MECE (mutually exclusive and collectively exhaustive) by construction, we are left with *(H1.h1)*, that is, with the conclusion that protocol differences exist and cannot be explained by rational-choice benchmarks. This achieves the goal of the present paper, as stated in the Introduction.

Exploratory analysis

Future research is required to propose and test alternative behavioral theories towards better organizing the existing experimental data including our own: standard controls provide no satisfactory explanation.

Interactive dictator games –which have been played in economic laboratories without being analyzed for what they are– are interesting interactions of their own and deserve further game-theoretic and experimental attention, in particular in relation with hybrid models of bounded rationality and altruism.[30] As an outlook for what kinds of research directions could be taken in the future, we briefly explore two popular alternative models (one of bounded rationality, and one where the act of giving itself, independent of consequences, generates utility), as well as a discussion of what in our opinion is perhaps the most promising candidate.

Level-k reasoning. The intuition given after stating Theorem 1 can also be used to sketch what kinds of predictions models of cognitive hierarchy and level-k reasoning (see Camerer et al. 2004, and Crawford et al. 2013 for a review) would produce. Specifically, as long as each player wants others to have less than himself/herself, then such preferences would drive mutual giving decisions down in the interactive game compared with unilateral giving decisions in the non-interactive setting. However, only if players are strategically perfectly rational ($k \to \infty$: iterative reasoning is applied ad infinitum) payments arrive at zero. Thus, cognitive hierarchy/level-k reasoning offers a potential explanation for the existence of intermediate payments under the interactive protocol. However, such theories predict that average payments from interactive treatments are lower than from the non-interactive treatments – which conflicts with our experimental observation of the reverse relation.

Warm glow. The theory of 'warm glow' (Andreoni, 1989, 1990) posits that the act of giving itself generates utility – regardless of what others do. It thus seems plausible that in the presence of warm glow the zero-giving equilibrium in an interactive setting would disappear. And, indeed, this is the case as shown in Proposition 5 in the Appendix A.8. Hence, warm glow offers an alternative explanation of non-extremal payments in the interactive treatment; in the non-interactive case, it is not possible to distinguish pure altruism from impure altruism (warm glow) without additional controls. Crucially, warm glow alone, because of its independence of strategic incentives, can also not explain the observed protocol differences, that is, why payments in the interactive setting are higher than in the non-interactive setting.

Beliefs. Since none of the 'standard' behavioral models discussed above was able to organize the data satisfactorily, we conducted a secondary study – exploratory rather than falsificatory in nature. Under various interactive treatments, a total of 522 subjects played a single dictator game each (with $p = 1$) and reported a belief concerning what they expected others' giving decisions to be in the same game, see Table 5 and Wehrli (2018) for experimental details. Treatments varied with respect to the lengths of the 'loops' that subjects belonged to, and we used the protocol variation to identify whether belief effects (for which interactive protocols as used in the literature do not control) have significant traces.[31] Indeed, we find that between 38 and 56 percent of the subjects give amounts exactly equal to their beliefs. This makes the alternative explanation plausible that a substantial part of the observed heterogeneity of giving decisions in interactive protocols, which thus far has been interpreted as heterogeneity in other-regarding preferences, stems instead from simpler (equality-oriented)

[30] While we should certainly not avoid interactive dictator games altogether in the future, we should think about the interest they may have in their own right because they might be closer to several 'real world' giving situations than the non-interactive dictator games with which those have been studied thus far; see also Footnote 3.

[31] We considered loop lengths of 2, 3, 5 or involving the entire session (recruited to be 50).

heuristics with giving heterogeneity driven by belief heterogeneity.[32] Beyond beliefs about others' giving behavior as a basis for best response, beliefs could also be essential for the understanding of how –descriptive or injunctive– norms govern behavior. Whether those can explain behavior (see e.g. Krupka and Weber, 2013 for an elicitation method relevant to our setting) is a promising avenue for future research. Certainly, more appropriate controls for the role of beliefs in interactive protocols are needed.

In sum, none of the models that are frequently used in this field of research can organize the data of our experiment, or indeed of prior experimental studies: Nash equilibrium predictions relying on common other-regarding preference models (CES, Fehr-Schmidt, Bolton-Ockenfels, Charness-Rabin) are way off; warm-glow giving cannot account for the observed protocol effects; cognitive hierarchy/level-k reasoning, if anything, suggest effects in the opposite directions. Future experimental and theoretical work is needed to organize the larger body of existing data (including ours). The inclusion of auxiliary controls such as concerning strategic sophistication and awareness, as well as for beliefs regarding others' giving propensities and norm perceptions might turn out to be important in that effort.

5. Conclusion

Reciprocal motivations where beliefs matter are the key constituents of many models used in behavioral economics.[33] Even so, the extensive experimental literature concerned with other-regarding preference estimation has not considered them – despite the fact that the actual games being played in many of these studies are interactive dictator games, where giving and receiving go hand in hand. From a game-theoretic perspective, those games are more similar to voluntary contribution games than to non-interactive dictator games.[34] Yet, studies based on interactive dictator games pursue a line of analysis as if the underlying decision context featured no strategic interaction whatsoever. The present paper makes an explicit comparison of non-interactive and interactive protocols both in terms of theory foundations and by means of experimental tests that are designed to falsify. Both lines of investigation find complementary evidence showing that the two protocols should not be used interchangeably.[35]

Specifically, our theoretical analysis of interactive dictator games has shown that under standard rationality assumptions, Nash equilibria can even feature extremal payments for all preference types – in our specific model, this happens generically. In particular, even for substantially altruistic players, the equilibrium can be zero giving. This stands in stark contrast to non-interactive dictator games where rational players would make intermediate payments (unless narrowly self-interested/selfless). Based purely on data of giving decisions without further controls regarding strategic awareness and beliefs about others' decisions, it is thus impossible to disentangle strategic sophistication from other-regarding preferences in an interactive settings: this can even mean that if someone makes consistent non-zero payments, then we are able to conclude that he/she is not completely selfish, but we cannot positively deduce how altruistic he/she is as we do not know his/her beliefs and cannot control for his/her level of strategic reasoning. If, instead, the observed payments are consistently zero, then we cannot know whether this is the result of pure selfishness or of full rationality (with perhaps substantial altruism). Thus, our theory results alone reveal a fundamental problem regarding the way interactive dictator games are usually interpreted in preferences estimation studies.

To complement our theoretical investigation, we conducted a pre-registered experiment to isolate the protocol difference behaviorally, using the original Andreoni and Miller (2002) framework with minimal modifications required for proper testing. The purpose of the experiment was a pre-registered 'clean' test of the protocol effect, and its results have revealed a significant difference in giving decisions between non-interactive and interactive protocols, thereby rejecting the interchangeability of non-interactive and interactive protocols also from an experimental perspective. More precisely, even under the –in our view implausible[36]– assumption that players have no strategic awareness or sophistication, our experimental results would still indicate a significant bias when using an interactive protocol instead of a non-interactive one. What is more, giving decisions in the interactive setting did not follow pure-strategy Nash equilibrium: in particular, intermediate giving continued to dominate, thereby adding serious attenuation to the assertion of 'rational' altruism in such settings. Other commonly used candidate explanations (warm glow, cognitive hierarchy) could not organize the data either. Exploring the data further and running additional experiments, we find that without additional auxiliary variables (measures of strategic sophistication and awareness) and controls (concerning beliefs regarding others' giving) behavior in interactive dictator games cannot satisfactorily be explained.

One is left wondering 'So what do we really learn from dictator game experiments?'. One thing that all dictator game experiments (including this study) have done time and again (regardless of the underlying protocol) is falsifying the ex-

[32] Further work is needed to disentangle this hypothesis from an alternative hypothesis according to which beliefs might not be truthfully reported but rather stated so as to 'justify' giving decisions.

[33] Reciprocity is crucial, for example, in many models of other-regarding preferences (Sobel, 2005) and in public goods games (Chaudhuri, 2011). The strategy method is one of the many techniques that is used to spell this out, and could be used in interactive dictator games in the future.

[34] Grech (2019) makes this connection explicit.

[35] A third variant of the dictator game that is often implemented –yet not analyzed here– is one where there is not actually any role duality, but instead uncertainty regarding whose decisions are implemented in the dictator's role. In a modified dictator game (including actions of surplus destruction) Iriberri and Rey-Biel (2011) find dramatic treatment differences.

[36] Note that in our experiment 43% of participants in the interactive setting indicated to have both understood and included the strategic nature of the game in their decision-making.

treme view that humans are narrowly self-interested all the time.[37] Yet, while falsification is one of the principal aims of experimentation –indeed the main aim of our own experiment– the hope is of course that we will be able to learn more from them than that in the future. Our study indicates that whether or not one is able to do so critically depends on seemingly subtle details of the experimental design – a general observation that other authors have mentioned before (see e.g. Levitt and List 2007). The specific case considered here is special in that it goes beyond the usual critique regarding framing effects: we find that an entirely different *game* is often implemented. Thus, the quote by Forsythe et al. at the beginning of this paper may be true in theory, but in practice the 'non-game' feature of the actual dictator game is given up in many experimental implementations. Social preference measurements obtained in such a way are therefore confounded, since beliefs and interactive considerations enter in hitherto uncontrolled ways, and valid alternative interpretations exist. We believe this to be a sensitive issue, as the other-regarding preference estimation literature is oftentimes concerned with value-laden differences among humans based on factors such as social class, gender, generation, nationality, etc. (see Footnote 4). Our hope with the present paper is to initiate and contribute to the development of better tools to control for potential confounds when conclusions of high societal relevance are at stake.

Acknowledgments

We thank Stefan Wehrli for his comments and help with the experiment. We also thank the editor Vince Crawford, the anonymous advisory editor, and two anonymous referees at Games and Economic Behavior for constructive criticism. Further, we thank Kurt Ackermann, Oliver Brägger, Stefano Duca, Dirk Engelmann, Matthew Jackson, Shachar Kariv, Caleb Koch, Michael Koch, Tatyana Kovalenko, Moritz Kraemer, Ryan Murphy, Matthew Polisson, Bary Pradelski, Henri Rästas, Alvin Roth, Larry Samuelson, Eran Shmaya, Shyam Sunder, Joachim Weimann and Peyton Young for help and stimulating discussions. We also profited from comments by Andreas Steingötter of the Seminar for Statistics at the ETH Zurich, by an anonymous reviewer of the Open Science Framework, and by participants at the Computational Social Science Seminar, the 2019 Swiss Theory Day in Basel, the 2018 Lisbon Meetings, the Behavioral Research Seminar at University of Bern, the 2018 Annual Conference of the Gesellschaft für experimentelle Wirtschaftsforschung, the 2018 Stony Brook International Conference on Game Theory and the 2017 Group Decision and Negotiation Conference. The submission evolution of this paper was frustrating, perhaps resulting in a more pointed exposition than we had originally intended. Sorry. Nax's position while at ETH Zurich was funded by the European Commission through the ERC Advanced Investigator Grant 'Momentum' (Grant No. 324247), while currently it is funded by an SNF Eccellenza Grant at UZH. Further, we are grateful to have received generous support from the Gesellschaft für experimentelle Wirtschaftsforschung through a 'Heinz Sauermann-Förderpreis' towards covering the costs of the experiment.

Appendix A. Proofs and additional mathematical results

A.1. Proof of Theorem 1

Best replies. From Eqs. ((5),(6)) (π_i being shorthand for $\mathcal{BR}_i(\pi_{-i})$), we compute

$$\begin{aligned}
&1-\pi_i+p\pi_{i-1}=a_i\sum_{j\neq i}w_j^{(i)}\left(1-\pi_j+p\pi_{j-1}\right)\\
\Leftrightarrow\;&1-\pi_i+p\pi_{i-1}=a_i-a_iw_{i-1}^{(i)}\pi_{i-1}+pa_iw_{i+1}^{(i)}\pi_i+a_i\sum_{j\neq i,i-1}(pw_{j+1}^{(i)}-w_j^{(i)})\pi_j\\
\Leftrightarrow\;&1-a_i=\left(1+pa_iw_{i+1}^{(i)}\right)\pi_i-\left(p+a_iw_{i-1}^{(i)}\right)\pi_{i-1}+a_i\sum_{j\neq i,i-1}\left(pw_{j+1}^{(i)}-w_j^{(i)}\right)\pi_j.
\end{aligned}$$

The following expression for the best reply follows:

$$\begin{aligned}
&\mathcal{BR}_i(\pi_{-i})\\
&=\min\left\{1,\max\left\{0,\left(1+pa_iw_{i+1}^{(i)}\right)^{-1}\left[1-a_i+\left(p+a_iw_{i-1}^{(i)}\right)\pi_{i-1}-a_i\sum_{j\neq i,i-1}\left(pw_{j+1}^{(i)}-w_j^{(i)}\right)\pi_j\right]\right\}\right\} \qquad (10)
\end{aligned}$$

What remains is to check the extremal cases that were excluded by (5). As mentioned in Subsection 3.2, these are in fact (almost) covered by (10).

Perfect egoist. If $c_i=0$, then $u_i(\pi_i,\pi_{-i})=1-\pi_i+p\pi_{-i}$, so that $\mathcal{BR}_i(\pi_{-i})\equiv 0$ regardless of ρ_i; this case is absorbed by allowing for $a_i=\infty$.

[37] Oechssler (2010) goes as far as suggesting that this conclusions is indeed all that can be learned from dictator games.

Perfect altruist. If $c_i = 1$, then $u_i(\pi_i, \pi_{-i}) = \frac{1}{N-1}\sum_{j\neq i}\left(1 - \pi_j + p\pi_{j-1}\right)$, so that $\mathcal{BR}_i(\pi_{-i}) \equiv 1$ regardless of ρ_i; this case is absorbed by allowing for $a_i = 0$.

Rawlsian. If $\rho_i \to -\infty$, we obtain a Leontief utility function $u_i(\pi_1, \ldots, \pi_N) = \min\{o_i, s_i\}$ and hence Player i strives for equality, i.e. $s_i = o_i$; consistent with $a_i = 1$, cf. Equation (6).

Efficiency optimizer. If $\rho_i = 1$, we distinguish three subcases:

(*a*) If $w_{i+1}^{(i)} pc_i > (1 - c_i)$, Player i's best reply is $\mathcal{BR}_i(\pi_{-i}) \equiv 1$, corresponding to $a_i = 0$.

(*b*) If $w_{i+1}^{(i)} pc_i < (1 - c_i)$, we have $\mathcal{BR}_i(\pi_{-i}) \equiv 0$ which amounts to setting $a_i = \infty$.

(*c*) If $w_{i+1}^{(i)} pc_i = (1 - c_i)$, Player i's utility is independent of his/her own payment, and thus any payment is a best reply. This is the only case that is not absorbed by a particular value of a_i and hence neither by Equation (10).

Proof of Item 1. If $\rho_i = 1$ and $w_{i+1}^{(i)} pc_i = (1 - c_i)$, then any payment may be a best reply, in particular $\pi_i = 1$. For $(\pi_1^*, \ldots, \pi_N^*) = (1, \ldots, 1)$ to be an equilibrium, in particular, it therefore suffices to show, cf. Equation (10), that

$$\left(1 + pa_i w_{i+1}^{(i)}\right)^{-1}\left[1 - a_i + \left(p + a_i w_{i-1}^{(i)}\right) - a_i \sum_{j\neq i,i-1}\left(p w_{j+1}^{(i)} - w_j^{(i)}\right)\right] \geq 1.$$

Using $\sum_{j\neq i} w_i^{(i)} = 1$ the left hand side simplifies to $\left(1 + pa_i w_{i+1}^{(i)}\right)^{-1}\left(1 + pa_i w_{i+1}^{(i)} + p(1 - a_i)\right)$, and hence the claim follows from $0 \leq a_i \leq 1$.

Proof of Item 2. It is clear that $(\pi_1^*, \ldots, \pi_N^*) = (0, \ldots, 0)$ is an equilibrium, because if $\pi_j = 0$ for all $j \neq i$, then all we need to show, cf. Equation (10), is

$$\left(1 + pa_i w_{i+1}^{(i)}\right)^{-1}(1 - a_i) \leq 0,$$

which holds by the assumption $1 \leq a_i$. □

A.2. Theorem 1 with arbitrary budgets and multipliers

If we allow players to have different budgets (endowments) $b_i > 0$ and multipliers of redistribution $p_i > 0$ $(i = 1, \ldots, N)$, Equation (3) generalizes to

$$u_i(\pi_1, \ldots, \pi_N) = \left[c_i\left(\sum_{j\neq i} w_j^{(i)}(b_j - \pi_j + p_{j-1}\pi_{j-1})\right)^{\rho_i} + (1 - c_i)(b_i - \pi_i + p_{i-1}\pi_{i-1})^{\rho_i}\right]^{\frac{1}{\rho_i}},$$

and likewise the expression for best replies, Equation (10), to

$$\begin{aligned}&\mathcal{BR}_i(\pi_{-i})\\ &= \min\left\{b_i, \max\left\{0,\right.\right.\\ &\qquad\left.\left.\left(1 + p_i a_i w_{i+1}^{(i)}\right)^{-1}\left[b_i - a_i\sum_{j\neq i} w_j^{(i)} b_j + \left(p_{i-1} + a_i w_{i-1}^{(i)}\right)\pi_{i-1} - a_i\sum_{j\neq i,i-1}\left(p_j w_{j+1}^{(i)} - w_j^{(i)}\right)\pi_j\right]\right\}\right\},\end{aligned}$$

where $a_i = \left(\frac{w_{i+1}^{(i)} p_i c_i}{1 - c_i}\right)^{\frac{1}{\rho_i - 1}}$. Then, $(\pi_1^*, \ldots, \pi_N^*) = (0, \ldots, 0)$ is an equilibrium if for all i

$$b_i \leq a_i \sum_{j\neq i} w_j^{(i)} b_j, \tag{11}$$

and $(\pi_1^*, \ldots, \pi_N^*) = (b_1, \ldots, b_N)$ is an equilibrium if for all i

$$p_{i-1} b_{i-1} \geq a_i \sum_{j\neq i} p_{j-1} w_j^{(i)} b_{j-1}. \tag{12}$$

A.3. Theorem 1 with uncertainty

Interactive implementations feature various degrees of uncertainty regarding the other player(s), as e.g. in Fisman et al. (2015b). These include incomplete information regarding preference types as well as regarding the budgets and multipliers of redistribution for other players. In this section we sketch how Theorem 1 translates to such settings. While we make no attempt at giving a full-fledged equilibrium analysis for the most general game here (as this would fill another paper and likely result in many different equilibria), we use this section to identify circumstances in which one can make use of the particular structure of the Equilibrium Conditions (11): it turns out that sizeable ranges of parameters can be pooled and lead to identical equilibria – closely connected to the bang-bang structure in Theorem 1. We use this fact to allow for the different types of uncertainty mentioned above 'as long as parameters stay within their pooling ranges'. We first introduce the shorthand notation for preference types,

$$t_i := (c_i, \rho_i, \{w_j^{(i)}\}_{j\neq i}) \in [0,1] \times (-\infty, 1] \times \Sigma^{N-2} := T,$$

(Σ^{N-2} being the $N-2$ dimensional standard simplex in $\mathbb{R}^{N-1}$) and allow for uncertainty expressed by a conditional distribution $f_i(t_{-i}, p_{-i}, b_{-i})$ regarding preference types $t_i \in T$, multipliers of redistribution $p_i \in [p_{\min}, p_{\max}] =: P$ and budgets $b_i \in [b_{\min}, b_{\max}] =: B$ for each Player i. Note that the game parameters p_i, b_i in this experimental context are drawn from a commonly known distribution with density h. By contrast, the t_i are a priori determined by subjective beliefs, i.e. $f_i(t_{-i}, p_{-i}, b_{-i}) = g_i(t_{-i})h(p_{-i}, b_{-i})$.

Each Player i then chooses a strategy $\pi_i^*(x_i)$, where $x_i := (t_i, p_i, b_i)$ is his/her full type. His/her expected utility satisfies

$$\mathbf{E}u_i(\pi_1^*(\cdot), \ldots, \pi_i, \ldots, \pi_N^*(\cdot)|x_i) \leq \mathbf{E}u_i(\pi_1^*(\cdot), \ldots, \pi_i^*(x_i), \ldots, \pi_N^*(\cdot)|x_i)$$

or, explicitly,

$$\begin{aligned}\int_{T^{N-1}\times P^{N-1}\times B^{N-1}} & u_i(\pi_1^*(x_1), \ldots, \pi_{i-1}^*(x_{i-1}), \pi_i, \pi_{i-1}^*(x_{i-1}), \ldots, \pi_N^*(x_N)) f_i(x_{-i}) \mathrm{d}^{n-1}x_{-i} \\ &\leq \int_{T^{N-1}\times P^{N-1}\times B^{N-1}} u_i(\pi_1^*(x_1), \ldots, \pi_N^*(x_N)) f_i(x_{-i}) \mathrm{d}^{n-1}x_{-i}, \end{aligned} \tag{13}$$

for every $i = 1, \ldots, N$, where $\pi_i \in [0, b_i]$, and $x_i \in T \times P \times B$. In general, choosing the best-responding strategy $\pi_i^*(x_i)$ constitutes a difficult optimization problem. However, extremal payments for all players and thus extremal equilibria survive under the following assumptions on the distribution $f_i(x_{-i})$ and in particular on the subjective beliefs $g_i(t_{-i})$. Namely, as long as everybody satisfies Condition (11) (Condition (12)) and in addition believes that this condition is satisfied for all i (i.e. other parameter configurations have weight zero), zero-giving (full-giving) is an equilibrium, in the sense that there are no profitable deviations from $\pi_i^*(x_i) \equiv 0$ for all i.[38] Indeed, in this case, we have for all such types

$$\begin{aligned}\int_{T^{N-1}\times P^{N-1}\times B^{N-1}} & u_i(0, \ldots, 0, \pi_i, 0, \ldots, 0) f_i(x_{-i}) \mathrm{d}^{n-1}x_{-i} \\ &= \int_{(11)\text{ satisfied}} u_i(0, \ldots, 0, \pi_i, 0, \ldots, 0) f_i(x_{-i}) \mathrm{d}^{n-1}x_{-i} \\ &\leq \int_{(11)\text{ satisfied}} u_i(0, \ldots, 0) f_i(x_{-i}) \mathrm{d}^{n-1}x_{-i}, \\ &\leq \int_{T^{N-1}\times P^{N-1}\times B^{N-1}} u_i(0, \ldots, 0) f_i(x_{-i}) \mathrm{d}^{n-1}x_{-i} \end{aligned}$$

(and similarly for full-giving). Thus, zero-giving (full-giving) is still an equilibrium despite considerable uncertainty in the game.[39]

[38] In case of a common prior regarding preference types t_i, this corresponds to the textbook definition of a Bayesian equilibrium. Note however, that common priors regarding the preference types t_i are not necessary to fulfill Condition (11).

[39] Bayesian equilibrium conditions represent a consistency check (similarly as for Nash equilibrium), which is why only first-order beliefs entered in our derivation. Higher-order beliefs, as for instance would be specified in case of common knowledge of (11) or (12), are needed to determine whether rational play resembles Bayesian equilibrium.

Example 1. Condition (11) can be rewritten as (we exclude corner cases for c_i, ρ_i for the sake of brevity)

$$\frac{c_i w_{i+1}^{(i)}}{1-c_i} \leq \frac{1}{p_i}\left(\frac{b_{i+1}}{b_i}\right)^{1-\rho_i} \qquad \text{for all } i=1,\dots,N. \tag{14}$$

In our own experiment, where $b_i = 1$ for all i, we thus recover, cf. Condition (7),

$$c_i \leq \frac{1}{1+w_i^{(i+1)} p_i},$$

so that, if at every individual level concerns satisfy *and* others' beliefs are believed to satisfy $c_i \leq \frac{1}{1+p_{\max}} = \frac{1}{3}$, then zero-giving is an equilibrium.

Example 2. The data in Fisman et al. (2015b) is such that $b_{\min} = 10$, $b_{\max} = 100$ and $p_i = \frac{b_i^{(i+1)}}{b_i}$, ($b_i^{(i+1)}$ being the maximally obtainable value of i's recipient) with the additional restriction that if $b_i \leq 50$, then $b_i^{(i+1)} \geq 50$. Thus, we have

$$\begin{aligned}
\frac{1}{p_i}\left(\frac{b_{i+1}}{b_i}\right)^{1-\rho_i} &\geq \frac{b_i^{\rho_i} b_{i+1}^{1-\rho_i}}{b_i^{(i+1)}} \\
&\geq \frac{b_i^{\rho_i}}{b_i^{(i+1)}} 10^{1-\rho_i} \\
&\geq \begin{cases} \frac{10^{\rho_i}}{100} 10^{1-\rho_i} = \frac{1}{10}, & \text{if } \rho_i \geq 0, \\ 100^{\rho_i - 1} 10^{1-\rho_i} = \frac{1}{10^{1-\rho_i}} & \text{if } \rho_i < 0, \end{cases} \qquad \text{for all } i=1,\dots,N.
\end{aligned}$$

If we assume $\rho_i \geq -1$, which is reasonable for many subjects, Condition (14) becomes

$$\frac{c_i w_{i+1}^{(i)}}{1-c_i} \leq \frac{1}{100} \Leftrightarrow c_i \leq \frac{1}{1+100 w_{i+1}^{(i)}} \qquad \text{for all } i=1,\dots,N, \tag{15}$$

which is admittedly less generic[40] but once again shows that interactive dictator games are very different from non-interactive ones in terms of rational-choice benchmarks.

A.4. Zero-giving under alternative other-regarding preferences

In this subsection, we briefly sketch how zero-giving equilibria arise in interactive 2-player dictator games under alternative other-regarding preference specifications. Several of these results readily generalize to $N > 2$-player dictator games for appropriate bracketing assumptions and preferences.

Let π_i denote player i's material payoff, and π_{-i} that of the other player. We consider the following preference models for player i.

Fehr-Schmidt (1999). The 2-player utility function reads

$$\begin{aligned}
u_i(\pi_i, \pi_{-i}) &= \begin{cases} s_i - \alpha_i(o_i - s_i); & \text{if } o_i \geq s_i \\ s_i - \beta_i(s_i - o_i); & \text{if } s_i > o_i. \end{cases} \\
&= \begin{cases} 1 - (1+\alpha_i(1+p))\,\pi_i + (p+\alpha_i(1+p))\,\pi_{-i}; & \text{if } \pi_i \geq \pi_{-i} \\ 1 - (1+\beta_i(1+p))\,\pi_{-i} + (p+\beta_i(1+p))\,\pi_i; & \text{if } \pi_{-i} > \pi_i \end{cases}
\end{aligned}$$

where $\beta_i \leq \alpha_i$ and $0 \leq \beta_i < 1$ parametrize the individual's inequity aversion. Without loss of generality we can assume that $\pi_i \geq \pi_{-i}$. But then $\pi_i = 0$ is the best reply for Player i and hence

[40] Note however that under broad bracketing with, say 51 subjects, Condition (15) becomes somewhat more realistic, namely $c_i \leq \frac{1}{3}$ for all i.

$$u_{-i}(\pi_{-i},\pi_i) = 1-(1+\alpha_{-i}(1+p))\pi_{-i}$$

from which it follows that $\pi_{-i}=0$ is indeed the best reply for the other player and hence $(\pi_1^*,\pi_2^*)=(0,0)$ is the unique Nash equilibrium in pure strategies.

Bolton-Ockenfels (2000). The 2-player utility function reads

$$u_i(\pi_i,\pi_{-i}) = v_i\left(s_i,\frac{s_i}{s_i+o_i}\right) = v_i\left(1-\pi_i+p\pi_{-i},\frac{1-\pi_i+p\pi_{-i}}{2+(p-1)(\pi_i+\pi_{-i})}\right)$$

where v_i is a twice differentiable function in both arguments with

$$\partial_1 v_i\left(s_i,\frac{s_i}{s_i+o_i}\right)\geq 0, \qquad \partial_1^2 v_i\left(s_i,\frac{s_i}{s_i+o_i}\right)\leq 0 \qquad \text{(narrow self-interest)}$$

$$\partial_2 v_i\left(s_i,1/2\right)=0, \qquad \partial_2^2 v_i\left(s_i,\frac{s_i}{s_i+o_i}\right)<0. \qquad \text{(comparative effect)}$$

For a given π_{-i} a best reply for Player i is determined by

$$\begin{aligned}0 &= \partial_{\pi_i} u_i(\pi_i,\pi_{-i})\\ &= \partial_1 v_i\left(s_i,\frac{s_i}{s_i+o_i}\right)\cdot(-1)+\partial_2 v_i\left(s_i,\frac{s_i}{s_i+o_i}\right)\cdot\left(-\frac{s_i p+o_i}{(s_i+o_i)^2}\right).\end{aligned} \tag{16}$$

Assuming $\partial_2 v_i\left(s_i,\frac{s_i}{s_i+o_i}\right)>0$, the right hand side of Equation (16) would be strictly negative and hence $\pi_i=0$ would be the best reply. In that case, $\frac{s_i}{s_i+o_i}=\frac{1+p\pi_{-i}}{2+(p-1)\pi_{-i}}>\frac{1}{2}$. However, by the comparative effect property we would also have $\frac{s_i}{s_i+o_i}\leq\frac{1}{2}$, a contradiction. It follows that $\partial_2 v_i\left(s_i,\frac{s_i}{s_i+o_i}\right)\leq 0$, which, again by the comparative effect property, is equivalent to $\frac{s_i}{s_i+o_i}\geq\frac{1}{2}$. In equilibrium, this holds for both players and therefore $\frac{s_i}{s_i+o_i}=\frac{s_{-i}}{s_{-i}+o_{-i}}=\frac{1}{2}$ resp. $\pi_i^*=\pi_{-i}^*$. As a consequence, $\partial_2 v_i\left(s_i,\frac{s_i}{s_i+o_i}\right)=\partial_2 v_i\left(s_i,\frac{1}{2}\right)=0$ in equilibrium and thus we have

$$\partial_{\pi_i} u_i(\pi_i^*,\pi_{-i}^*) = -\partial_1 v_i\left(1+(p-1)\pi_i^*,\frac{1}{2}\right).$$

Hence, in the generic case where $\partial_1 v_i\left(1,\frac{1}{2}\right)>0$ for at least one player, $(\pi_i^*,\pi_{-i}^*)=(0,0)$ is indeed the unique equilibrium. If $\partial_1 v_i\left(1,\frac{1}{2}\right)=0$ for both players, then by the narrow self-interest property, the function $\partial_1 v_i\left(\cdot,\frac{1}{2}\right)$ vanishes identically, that is, as long as payoffs are equally distributed, the players value all outcomes equally – they are Rawlsian. In particular, any (π_i^*,π_{-i}^*) with $\pi_i^*=\pi_{-i}^*$ is then an equilibrium.

Charness-Rabin (2002). The 2-player utility function reads[41]

$$\begin{aligned}u_i(\pi_i,\pi_{-i}) &= \begin{cases}(1-\sigma_i)s_i+\sigma_i o_i; & \text{if } o_i\geq s_i\\ (1-\rho_i)s_i+\rho_i o_i; & \text{if } s_i>o_i\end{cases}\\ &= \begin{cases}1+(p\sigma_i-(1-\sigma_i))\pi_i+(p(1-\sigma_i)-\sigma_i)\pi_{-i}; & \text{if } \pi_i\geq\pi_{-i}\\ 1+(p\rho_i-(1-\rho_i))\pi_i+(p(1-\rho_i)-\rho_i)\pi_{-i}; & \text{if } \pi_{-i}>\pi_i\end{cases}\end{aligned}$$

where $\rho\geq\sigma$ parametrize concerns for *self* and aggregated payoffs. Without loss of generality we may assume that $\pi_i\geq\pi_{-i}$ so that

$$u_i(\pi_i,\pi_{-i}) = 1+(p\sigma_i-(1-\sigma_i))\pi_i+(p(1-\sigma_i)-\sigma_i)\pi_{-i}.$$

Then $\pi_i=0$ is a best reply if and only if $p\sigma_i<1-\sigma_i$ or, equivalently, $\sigma_i<1/(1+p)$, regardless of the value of ρ_i. Intuitively, this requires that a player must be sufficiently selfish if his/her opponent obtains at least as a high a payoff. If both players satisfy this condition, the unique equilibrium is $(\pi_i^*,\pi_{-i}^*)=(0,0)$.

[41] Note that we focus on the outcome-oriented part of the utility and omit the reciprocity component here due to the one-shot nature of our situation.

A.5. Proof of Theorem 2

Proof of Item 1: narrow bracketing. We argue by contradiction and assume that $\pi^*_{i_0} < 1$ is the lowest of all payments. Using Equation (10) we then obtain

$$\begin{aligned}\pi^*_{i_0} &\geq \left(1 + pa_{i_0}\right)^{-1}\left[1 - a_{i_0} + p\pi^*_{i_0-1} + a_{i_0}\pi^*_{i_0+1}\right] \\ &\geq \left(1 + pa_{i_0}\right)^{-1}\left[1 - a_{i_0} + p\pi^*_{i_0} + a_{i_0}\pi^*_{i_0}\right],\end{aligned}$$

which results in

$$\pi^*_{i_0}\left(1 + pa_{i_0} - p - a_{i_0}\right) \geq 1 - a_{i_0} > 0.$$

Using $1 + pa_{i_0} - p - a_{i_0} = (1-p)(1-a_{i_0})$ and $a_{i_0} < 1$ this inequality simplifies to

$$\pi^*_{i_0}(1-p) \geq 1, \tag{17}$$

which is impossible given $\pi^*_{i_0} < 1$ and $p > 0$.

Proof of Item 1: broad bracketing. We argue by contradiction and distinguish two cases.
Case 1 ($p \geq 1$). Let $\pi^*_{i_1-1}$ be the highest of all payments and assume that $\pi^*_{i_1-1} < 1$. It follows that $\pi^*_{i_1} < 1$ and thus with Equation (10),

$$\begin{aligned}\pi^*_{i_1-1} &\geq \pi^*_{i_1} \\ &\geq \left(1 + \frac{a_{i_1}p}{N-1}\right)^{-1}\left[1 - a_{i_1} + \left(p + \frac{a_{i_1}}{N-1}\right)\pi^*_{i_1-1} - \frac{a_{i_1}(p-1)}{N-1}\sum_{j\neq i_1, i_1-1}\pi^*_j\right] \\ &\geq \left(1 + \frac{a_{i_1}p}{N-1}\right)^{-1}\left[1 - a_{i_1} + \left(p + \frac{a_{i_1}}{N-1} - \frac{a_{i_1}(p-1)(N-2)}{N-1}\right)\pi^*_{i_1-1}\right].\end{aligned}$$

This can be rewritten as

$$a_{i_1} - 1 \geq \left(p + \frac{a_{i_1}}{N-1} - \frac{a_{i_1}(p-1)(N-2)}{N-1} - 1 - \frac{a_{i_1}p}{N-1}\right)\pi^*_{i_1-1}, \tag{18}$$

which simplifies to

$$a_{i_1-1} - 1 \geq (a_{i_1-1} - 1)(1-p)\pi^*_{i_1-1},$$

and since $a_{i_1-1} < 1$ we obtain $1 \leq (1-p)\pi^*_{i_1-1}$, a contradiction; thus $\pi^*_{i_1-1} = 1$. Assume that $\pi^*_{i_1} < 1$ holds, given that $\pi^*_{i_1-1} = 1$, we obtain

$$1 > \pi^*_{i_1} \geq \left(1 + \frac{a_{i_1}p}{N-1}\right)^{-1}\left[1 - a_{i_1} + p + \frac{a_{i_1}}{N-1} - \frac{a_{i_1}(p-1)(N-2)}{N-1}\right]$$

which simplifies to $a_{i_1} > 1$, contradicting the assumption. Proceeding through the entire loop of players, we see that $(\pi^*_1, \ldots, \pi^*_N) = (1, \ldots, 1)$ is indeed the only possible equilibrium.
Case 2 ($0 < p < 1$). Let $\pi^*_{i_2}$ be the lowest of all payments. If $\pi^*_{i_2} < 1$, then it follows that

$$\begin{aligned}\pi^*_{i_2} &\geq \left(1 + \frac{a_{i_2}p}{N-1}\right)^{-1}\left[1 - a_{i_2} + \left(p + \frac{a_{i_2}}{N-1}\right)\pi^*_{i_2-1} - \frac{a_{i_2}(p-1)}{N-1}\sum_{j\neq i_2, i_2-1}\pi^*_j\right] \\ &\geq \left(1 + \frac{a_{i_2}p}{N-1}\right)^{-1}\left[1 - a_{i_2} + \left(p + \frac{a_{i_2}}{N-1} - \frac{a_{i_2}(p-1)(N-2)}{N-1}\right)\pi^*_{i_2}\right]\end{aligned}$$

and thus $1 \leq (1-p)\pi^*_{i_2}$ by the same reasoning as in Case 1 above. This however contradicts $p < 1$ and hence $\pi^*_{i_2} = 1$. By the minimality of $\pi^*_{i_2}$, it follows that $\pi^*_i = 1$ for all i.

Proof of Item 2: narrow bracketing. We assume that the highest of all payments, $\pi^*_{i_3}$, satisfies $\pi^*_{i_3} > 0$ and show that this leads to a contradiction. Indeed, using Equation (10), we have

$$\begin{aligned}\pi^*_{i_3} &\leq \left(1 + pa_{i_3}\right)^{-1}\left[1 - a_{i_3} + p\pi^*_{i_3-1} + a_{i_3}\pi^*_{i_3+1}\right]\\ &\leq \left(1 + pa_{i_3}\right)^{-1}\left[1 - a_{i_3} + p\pi^*_{i_3} + a_{i_3}\pi^*_{i_3}\right]\end{aligned}$$

Using $a_{i_3} > 1$ and the same transformations as in the proof of Item 1 we arrive again (cf. Inequality (17)) at

$$\pi^*_{i_3}(1-p) \geq 1,$$

which contradicts the assumption $p > 0$.

Proof of Item 2: broad bracketing. We argue by contradiction and distinguish again two cases.
Case 1 ($p \geq 1$). We assume that there is an equilibrium with at least one player, say i_4, such that $\pi^*_{i_4} > 0$. We first show that it then holds that $\pi^*_i > 0$ for all $i = 1, \ldots, N$. To see why that is, note that then Player i_4's best reply in equilibrium, see Equation (10), implies

$$0 < 1 - a_{i_4} + \left(p + \frac{a_{i_4}}{N-1}\right)\pi^*_{i_4-1} - \frac{a_{i_4}(p-1)}{N-1}\sum_{j\neq i_4, i_4-1}\pi^*_j,$$

which can only hold if $\pi^*_{i_4-1} > 0$, since $1 - a_{i_4} < 0$ and $\frac{a_{i_4}(p-1)}{N-1}\sum_{j\neq i_4, i_4+1}\pi^*_j \geq 0$. Applying this argument recursively shows that indeed $\pi^*_i > 0$ for all i. It follows that if at least one player satisfies $\pi^*_i = 0$ then there is nothing to show. We therefore assume without loss of generality that $\pi^*_{i_4-1} > 0$ is the lowest of all payments. Since then $\pi^*_{i_4} > 0$, we have

$$\begin{aligned}\pi^*_{i_4-1} &\leq \pi^*_{i_4}\\ &\leq \left(1 + \frac{a_{i_4}p}{N-1}\right)^{-1}\left[1 - a_{i_4} + \left(p + \frac{a_{i_4}}{N-1}\right)\pi^*_{i_4-1} - \frac{a_{i_4}(p-1)}{N-1}\sum_{j\neq i_4, i_4-1}\pi^*_j\right]\\ &\leq \left(1 + \frac{a_{i_4}p}{N-1}\right)^{-1}\left[1 - a_{i_4} + \left(p + \frac{a_{i_4}}{N-1} - \frac{a_{i_4}(p-1)(N-2)}{N-1}\right)\pi^*_{i_4-1}\right].\end{aligned}$$

Using $a_{i_4} > 1$ this simplifies to, cf. Equation (18),

$$1 \leq (1-p)\pi^*_{i_4-1}.$$

This, however, is impossible as $p > 0$ and $0 \leq \pi^*_{i_4-1} \leq 1$.
Case 2 ($0 < p < 1$). We argue again by contradiction. Let $\pi^*_{i_5}$ be the maximum of all payments. If $\pi^*_{i_5} = 0$, there is nothing to show. If $\pi^*_{i_5} > 0$, we have

$$\begin{aligned}\pi^*_{i_5} &\leq \left(1 + \frac{a_{i_5}p}{N-1}\right)^{-1}\left[1 - a_{i_5} + \left(p + \frac{a_{i_5}}{N-1}\right)\pi^*_{i_5-1} - \frac{a_{i_5}(p-1)}{N-1}\sum_{j\neq i_5, i_5-1}\pi^*_j\right]\\ &\leq \left(1 + \frac{a_{i_5}p}{N-1}\right)^{-1}\left[1 - a_{i_5} + \left(p + \frac{a_{i_5}}{N-1} - \frac{a_{i_5}(p-1)(N-2)}{N-1}\right)\pi^*_{i_5}\right]\end{aligned}$$

which, cf. Equation (18) again, is equivalent to $1 \leq (1-p)\pi^*_{i_5}$ and thus we get again a contradiction. □

A.6. Proof of Proposition 3

Best replies. In the 2-player case, the formula for the best replies, Equation (10) simplifies to

$$\mathcal{BR}_i(\pi_{-i}) = \begin{cases} 0 & \text{if } \pi_{-i} \in [0, \frac{a_i - 1}{a_i + p}],\\ \frac{1 - a_i + (a_i + p)\pi_{-i}}{1 + pa_i} & \text{if } \pi_{-i} \in (\frac{a_i - 1}{a_i + p}, \frac{a_i(1+p)}{a_i + p}),\\ 1 & \text{if } \pi_{-i} \in [\frac{a_i(1+p)}{a_i + p}, 1],\end{cases} \tag{19}$$

assuming the same provisions for the a_i as in the N-player case.

Proof of Item 1. We first show that for $a_i a_{-i} \neq 1$, all pure Nash equilibria are located on the boundary of the unit square. Suppose not, i.e. $0 < \pi_i < 1$ for $i = 1, 2$. Then the formula for best replies yields the fixed point condition for an interior equilibrium,

$$\begin{pmatrix} 1 + pa_i & -(a_i + p) \\ -(a_{-i} + p) & 1 + pa_{-i} \end{pmatrix} \begin{pmatrix} \pi_i^* \\ \pi_{-i}^* \end{pmatrix} = \begin{pmatrix} 1 - a_i \\ 1 - a_{-i} \end{pmatrix}. \tag{20}$$

If $p \neq 1$ the determinant of the matrix on the left hand side is non-zero,

$$(1 + pa_i)(1 + pa_{-i}) - (a_i + p)(a_{-i} + p) = (p^2 - 1)(a_i a_{-i} - 1) \neq 0,$$

and it is readily seen that the unique solution of Equation (20) is then given by $\pi_i^* = \pi_{-i}^* = \frac{1}{1-p}$, which does not lie on the unit square $[0, 1]^2$ since $p > 0$. If $p = 1$, then (20) can be rewritten as

$$\begin{pmatrix} \pi_i^* - \pi_{-i}^* \\ \pi_{-i}^* - \pi_i^* \end{pmatrix} = \begin{pmatrix} \frac{1-a_i}{1+a_i} \\ \frac{1-a_{-i}}{1+a_{-i}} \end{pmatrix}, \tag{21}$$

which only has a solution if

$$\frac{1 - a_i}{1 + a_i} = \frac{a_{-i} - 1}{1 + a_{-i}} \Leftrightarrow a_i a_{-i} = 1,$$

and thus we arrive again at a contradiction.

We now construct the Nash equilibria on the boundary of $[0, 1]^2$ under the condition $a_i a_{-i} \neq 1$. Since the conditions for Item 1.*i* and 1.*iii* are slightly relaxed compared to those in Theorem 2, we give a separate proof here. However, we restrict ourselves to Items 1.*i* and 1.*ii* in Theorem 3. Items 1.*iii* and 1.*iv* are proved analogously.

Item 1.i ($a_i a_{-i} > 1$ and $a_{-i}, a_i \geq 1$). If $\pi_{-i}^* = 0$ then $\mathcal{BR}_i(0) = 0$ because $a_i \geq 1$. Since moreover $\mathcal{BR}_{-i}(0) = 0$ due to $a_{-i} \geq 1$ we conclude that $(\pi_i^*, \pi_{-i}^*) = (0, 0)$ is indeed an equilibrium. If $\pi_{-i}^* = 1$, then $\mathcal{BR}_i(1) = \frac{1+p}{1+pa_i}$ because $a_i \geq 1$. However, $\mathcal{BR}_{-i}(\frac{1+p}{1+pa_i}) = 1$ if only if

$$\begin{aligned} &\frac{a_{-i}(1 + p)}{a_{-i} + p} \leq \frac{1 + p}{1 + pa_i} \\ \Leftrightarrow\ & a_{-i} + pa_i a_{-i} \leq a_{-i} + p \\ \Leftrightarrow\ & a_i a_{-i} \leq 1, \end{aligned} \tag{22}$$

a contradiction. Thus $(\pi_i^*, \pi_{-i}^*) = (0, 0)$ is the only Nash equilibrium where $\pi_{-i}^* \in \{0, 1\}$, and by symmetry it is in fact the only equilibrium.

Item 1.ii ($a_i a_{-i} > 1$ and $a_{-i} > 1 > a_i$). If $\pi_{-i}^* = 0$ then $\mathcal{BR}_i(0) = \frac{1-a_i}{1+pa_i}$ because $a_i < 1$. Moreover, $\mathcal{BR}_{-i}(\frac{1-a_i}{1+pa_i}) = 0$ because

$$\begin{aligned} &\frac{1 - a_i}{1 + pa_i} \leq \frac{a_{-i} - 1}{a_{-i} + p} \\ \Leftrightarrow\ & 1 \leq a_i a_{-i} \end{aligned}$$

which is true by assumption and thus $(\pi_i^*, \pi_{-i}^*) = (\frac{1-a_i}{1+pa_i}, 0)$ is indeed an equilibrium. There are no other equilibria:

(*a*) If $\pi_{-i}^* = 1$, then $\mathcal{BR}_i(1) = 1$ because $a_i < 1$; yet $\mathcal{BR}_{-i}(1) = \frac{1+p}{1+pa_{-i}} < 1$ since $a_{-i} > 1$, a contradiction.

(*b*) If $\pi_i^* = 0$, then $\mathcal{BR}_{-i}(0) = 0$ because $a_{-i} > 1$; yet $\mathcal{BR}_i(0) = \frac{1-a_i}{1+pa_i} > 0$ since $a_i < 1$, a contradiction.

(*c*) If $\pi_i^* = 1$ then $\mathcal{BR}_{-i}(1) = \frac{1+p}{1+pa_{-i}} < 1$ because $a_{-i} > 1$; yet $\mathcal{BR}_i(\frac{1+p}{1+pa_{-i}}) < 1$: to see this, assume $\mathcal{BR}_i(\frac{1+p}{1+pa_{-i}}) = 1$. But then $\frac{1+p}{1+pa_{-i}} \geq \frac{a_i(1+p)}{a_i+p}$ which is equivalent to $1 \geq a_i a_{-i}$, a contradiction.

Proof of Item 2. Now let $a_i a_{-i} = 1$ and without loss of generality $a_{-i} \geq 1 \geq a_i$. We start by determining the Nash equilibria on the boundary. To that end, let us first look at the case $a_{-i} = a_i = 1$. With the same arguments used to prove Item 1.*i* above –considering that (22) is now satisfied– we conclude that $(\pi_i^*, \pi_{-i}^*) = (0, 0)$ and $(\pi_i^*, \pi_{-i}^*) = (1, 1)$ are the only equilibria on the boundary of $[0, 1]^2$. Second, we consider the case $a_{-i} > 1 > a_i$, and apply the arguments used to prove

Item 1.*ii* above –considering that (*c*) no longer yields a contradiction– to conclude that $(\pi_i^*, \pi_{-i}^*) = (\frac{1-a_i}{1+pa_i}, 0)$ is the only equilibrium with $\pi_{-i}^* \in \{0,1\}$ and $(\pi_i^*, \pi_{-i}^*) = (1, \frac{1+p}{1+pa_{-i}})$ is the only equilibrium with $\pi_i^* \in \{0,1\}$. Finally, we turn to interior Nash equilibria. Multiplying the second line of (20) by $-a_i$, using $a_i a_{-i} = 1$ we obtain

$$\begin{pmatrix} 1+pa_i & -(a_i+p) \\ 1+pa_i & -(a_i+p) \end{pmatrix} \begin{pmatrix} \pi_i^* \\ \pi_{-i}^* \end{pmatrix} = \begin{pmatrix} 1-a_i \\ 1-a_i \end{pmatrix}. \tag{23}$$

It is straightforward to see that this system has the admissible solutions

$$\left\{ (\pi_i, \pi_{-i}^*) = \left(\frac{1-a_i+(a_i+p)t}{1+pa_i}, t \right) \middle| 0 \le t \le \frac{a_i+pa_i}{a_i+p} = \frac{1+p}{1+pa_{-i}} \right\}.$$

Note that we recover the above determined boundary solution for extremal values of t.

Proof of Item 3. This follows from the fact that $\mathcal{BR}_i(\pi_{-i}) \equiv 1$ and $\mathcal{BR}_{-i}(\pi_i) \equiv 0$ in this case.

Proof of Item 4. This follows directly from the fact the Player i's utility is independent of π_i if $\rho_i = 1$ and that $c_i p = 1 - c_i$. □

A.7. Probability of being in a loop of given size

With the following proposition, we can show that the loop size requirement for a zero-giving equilibrium under broad bracketing, as given implicitly by Condition (9), holds with high probability in a typical experimental session.

Proposition 4. *Assume that $M \ge 2$ players are randomly matched in such a manner that every player is a dictator of exactly one player and a recipient of exactly one player. All matches between players are equally probable and subject to the constraint that every so created loop of players has at least size L. Then, the probability that a given player is in a loop with more than N_0 players is given by*

$$P(M, L, N_0) = \frac{M - N_0}{M - L + 1}.$$

Usually, the multiplier of redistribution satisfies $p \le 4$ and other-regarding concerns are bounded by the social norm $c_{\max} = 1/2$. Condition (9) for the zero-giving equilibrium thus holds if

$$N > \frac{pc_{\max}}{1 - c_{\max}} + 1 = \frac{4 \cdot 1/2}{1 - 1/2} + 1 = 5 =: N_0.$$

A typical experimental session involves about $M \ge 20$ subjects. In addition, it is usually not allowed for two participants to be each others' dictators and recipients at the same time, that is, $L = 3$. Thus, using Proposition 4, the probability for a given player to be in such a loop of zero giving is substantial, namely at least

$$P(M, L, N_0) = \frac{M - N_0}{M - L + 1} = \frac{20 - 5}{20 - 3 + 1} \approx 0.83.$$

Proof of Proposition 4. Let A_{N_0} be the event that a given Player i^* is in a loop of size bigger than N_0 ($\ge L$) and let B_k denote the event that Player i^* is *not* in a loop of length k. The probability of A_N is given by

$$\begin{aligned} \mathbf{P}[A_{N_0}] = \mathbf{P}\left[\bigcap_{k=L}^{N_0} B_k \right] &= \prod_{k=L}^{N_0} \mathbf{P}\left[B_k \middle| \bigcap_{l=L}^{k-1} B_l \right] \\ &= \prod_{k=L}^{N_0} \left(\frac{M-k}{M-k+1} \right) = \frac{M - N_0}{M - L + 1}, \end{aligned}$$

where the first equality holds since loops of length smaller than L are excluded, and the second because of the chain rule.

As for the third equality, it helps to think of the network structure being created by successively adding directed links between players, starting with Player i^*, until every player has exactly one inward and one outward going link. Assuming that Player i^* is not in any loop of size $l \le k-1$, he/she will eventually be the starting point of a directed chain of length

$k-1$ in this network creation process. The probability that this chain will not reconnect to a loop equals the probability that it will not reconnect to Player i^* but continue to one of the other available $M-(k-1)-1=M-k$ players, so that indeed

$$\mathbf{P}\left[B_k \middle| \bigcap_{l=L}^{k-1} B_l\right] = \frac{M-k}{M-k+1}.$$

The fourth equality above is obtained by omitting equal numerators and denominators in the product. □

A.8. Interactive dictator games in presence of warm glow

Here, we prove the absence of zero-giving in equilibrium for interactive dictator games if players' preferences include a warm-glow component. We replace the Definition (3) with the utility function (for the case of $\rho_i \neq 0$) which reads as

$$u_i^{\mathrm{WG}}(\pi_i, \pi_{-i}) := \left[c_i \left(\sum_{j\neq i} w_j^{(i)} (1-\pi_j + p\pi_{j-1}) \right)^{\rho_i} + d_i\,(1-\pi_i + p\pi_{-i})^{\rho_i} + g_i \pi_i^{\rho_i} \right]^{\frac{1}{\rho_i}},$$

where c_i, d_i denote Player i's concern for *self* and *other* respectively as before, and g_i measures 'warm glow', assuming that $c_i + d_i + g_i = 1$. As before, for the special case of $\rho_i = 0$, the same function is defined by a Cobb-Douglas utility

$$u_i^{\mathrm{WG,CD}}(\pi_i, \pi_{-i}) := \left(\sum_{j\neq i} w_j^{(i)} (1-\pi_j + p\pi_{j-1}) \right)^{c_i} (1-\pi_i + p\pi_{-i})^{d_i}\, \pi_i^{g_i}.$$

Proposition 5. *In the interactive dictator game, $(\pi_1, \ldots, \pi_N) = (0, \ldots, 0)$ cannot be a Nash equilibrium if at least one of the players satisfies non-zero warm glow: $g_i > 0$ and $\rho_i < 1$.*[42]

Proof. We compute

$$\partial_{\pi_i} u_i^{\mathrm{WG}}(\pi_i, 0, \ldots, 0) = \left(u_i^{\mathrm{WG}}(\pi_i, 0, \ldots, 0)\right)^{1-\rho_i} \left(w_{i+1}^{(i)} p c_i \left(1 + w_{i+1}^{(i)} p \pi_i\right)^{\rho_i - 1} - d_i (1-\pi_i)^{\rho_i - 1} + g_i \pi_i^{\rho_i - 1} \right).$$

If $g_i > 0$, then this last expression is positive (in fact, infinitely large) for $\pi_i \searrow 0$ since $\rho_i < 1$. Consequently, $\pi_i = 0$ cannot be a best reply to $\pi_{-i} = (0, \ldots, 0)$, which proves the claim for $\rho_i \neq 0$. The proof in the Cobb-Douglas case follows along the same lines. □

Appendix B. Theoretical implications of experimental hypotheses

Our three hypotheses *(H0), (H1.h0), (H1.h1)* are MECE. Here, we discuss the logical consequences of all potential outcomes of our test procedure in detail.

Case 1: Cannot reject (H0). Protocol differences are not significant, implications are inconclusive in a Popperian sense as nothing is falsified and we stop testing.

Notwithstanding, if we were to interpret this outcome to mean that no protocol differences exist, two causes are possible. Either subjects could in fact show different behavior under the two protocols on an individual basis which simply does not show in the (identical) overall distributions. As per definition of this possibility, preference estimation techniques would then be inapplicable.[43] Alternatively, each individual subject might indeed make identical decisions under both protocols, because he/she perceives the interactive protocol like the non-interactive one. Preference estimation would then *a priori* be valid even under an interactive design.

However, only two underlying reasons for such strategic unawareness are conceivable, both of which are problematic. For one, it could be that it is a general trait of humans playing dictator games. This conclusion would hint at a model of an economic agent incapable of any 'cognitive empathy' (as mentioned in the Introduction) – a concept, that would conflict

[42] In the special case of an efficiency optimizer ($\rho_i = 1$), this condition has to be modified to $g_i > d_i - c_i p w_{i+1}^{(i)}$ instead.

[43] Even with alternative experimental designs it is however difficult, if not impossible, to investigate the specifics of such an individual difference in behavior (see Footnote 19).

with a large body of research in behavioral game theory, which shows that most subjects are (at least boundedly) able to reason strategically in almost all other experimental games. The other reason would be that strategic unawareness is the result of instructions 'framing' otherwise strategic subjects in a way that clouds the actual rules of interactive dictator games – this, alas, would not constitute best practice in experimental economics, because we, as analysts of these games, could never be confident to what extent this framing was successful for all or some of the subjects.[44]

Case 2: Reject (H0). Protocol differences are significant, and we can conclude that strategic incentives have an effect.

This could be a consequence of strategic reasoning by individuals in a game-theoretic sense – in which case non-interactive preference estimation would be inappropriate and in many cases even ambiguous by virtue of our theoretical results, even if subjects are not perfectly strategic.[45] Alternatively, again upholding the extreme assumption that subjects are strategically entirely unaware, the rejection of *(H0)* might result from a some different (stochastic, socio-psychologic, etc.) unknown effect, in which case interactive preference estimation would *a priori* still be applicable. However, in addition to the problems mentioned above (in the case where we cannot reject *(H0)*), such measurements would also be subject to a significant bias compared to a standard non-interactive preference estimation.

In case of rejection of (H0), we have two cases.

Case 2.1: Cannot reject (H0.h0). Rational-choice benchmarks may serve as functioning behavioral theory explaining giving in interactive dictator games – again only in the Popperian sense that our experiment gives no reason for rejection.

Case 2.2: Reject (H0.h0). Observed differences deviate substantially from rational-choice benchmarks implying that those do not explain behavior in the interactive dictator game that we implemented.

In conclusion, even before collecting any experimental data, the above scenario analysis –based on the inextricableness of other-regarding motives and strategic considerations that was discovered on pure theoretical grounds– suggests that existing studies that are based on interactive protocols cannot exclude strategic confounds that may have altered their conclusions.

Appendix C. Additional statistical analyses and data

Table 3
p-Values of Non-parametric Tests for All Multipliers of Redistribution. We report the p-values of tests regarding all decisions, averages, and p-dependent decisions; Specifically, for differences in distribution using Kolmogorov-Smirnov (KS) tests and Mann-Whitney-Wilcoxon (MWW) tests, and for equality of variance using Levene and Brown-Forsythe (BF) (two-sided) tests. Stars indicate significance at the 95% confidence level. Grey-shaded p-values for MWW, Levene and BF highlight that KS indicated significant distributional differences.

Price of redistr. (p)	KS	MWW	Levene	BF	N. Obs.
All	< 0.001*	<0.001*	0.928	0.965	8'240
Average	0.055	< 0.001*	0.380	0.757	412
0.1	0.082	0.072	0.586	0.651	412
0.2	0.016*	0.007*	0.546	0.716	412
0.3	0.024*	0.009*	0.946	0.879	412
0.4	0.016*	0.007*	0.559	0.642	412
0.5	0.082	0.045*	0.775	0.915	412
0.6	0.016*	0.003*	0.621	0.606	412
0.7	0.024*	0.002*	0.738	0.895	412
0.8	0.038*	0.021*	0.968	0.768	412
0.9	0.038*	0.012*	0.692	0.669	412
1.0	0.163	0.064	0.216	0.364	412
1.1	0.055	0.017*	0.865	0.809	412
1.2	0.191	0.068	0.722	0.990	412
1.3	0.006*	0.002*	0.824	0.565	412
1.4	0.299	0.145	0.324	0.499	412
1.5	0.266	0.040*	0.804	0.907	412
1.6	0.096	0.051	0.319	0.302	412
1.7	0.122	0.026*	0.766	0.638	412
1.8	0.221	0.068	0.805	0.964	412
1.9	0.116	0.026*	0.538	0.521	412
2.0	0.144	0.032*	0.434	0.332	412

* Stands for $p < 0.05$.

[44] Recall that Case 1 is suggested by the findings of Korenok et al. (2013).
[45] Implications of k-level reasoning will be sketched below in the discussion of our experimental results.

Table 4
OLS of Giving Decisions (Exploratory Regression). Estimates of OLS regressions with individual-level clustering. Parentheses indicate [robust standard errors] and (p-values).

	Giving	
	(Reduced)	(Controls)
Interactive	65.767*	52.337
	[31.464]	[31.138]
	(0.037)	(0.094)
Inter.$\times p$	−6.045*	−6.045*
	[1.291]	[1.292]
	(<0.001)	(<0.001)
Non-int.$\times p$	−4.346*	−4.346*
	[1.279]	[1.280]
	(0.001)	(0.001)
Order		−0.077
		[0.277]
		(0.781)
Female		29.705
		[21.473]
		(0.167)
Age		−0.349
		[0.968]
		(0.718)
Self-centrism		−107.841*
		[34.069]
		(0.002)
Narrow bracketing		36.395
		[24.153]
		(0.133)
Interactive play		26.383
		[29.407]
		(0.370)
Constant	310.987*	298.047*
	[22.479]	[45.226]
	(<0.001)	(<0.001)
Root MSE	266.7	263.2
Number observations	8'240	8'240
Number individuals	412	412

* Stands for $p < 0.05$.

Table 5
Percentage of Giving Corresponding 1-1 with Beliefs Concerning Others' Giving.

Loop size	Percentage of 1-1 correspondences	Observations	Instructions
2	56%	100	You will be paired with another participant of this study. You split a dollar. The other receives what you transfer. The other also splits a dollar, and you receive what the other transfers to you.
3	38%	108	You will be paired with two other participants of this study. The three of you will form a ring. Each of you splits a dollar. Your next neighbor to the right receives what you transfer. Your neighbor to the left receives what your neighbor to the right transfers. You receive what your neighbor to the left has transferred.
5	43%	105	You will be paired with four other participants of this study. The five of you will form a ring. Each of you splits a dollar. Your next neighbor to the right receives what you transfer. And so it continues along the ring. You receive what your neighbor to the left has transferred.
entire session[a]	40%	209	You split a dollar. The amount shared by you will be transferred to a randomly paired MTurk worker who also participates in this study. Note that all participants face the same decision, and that you will be the recipient of another randomly paired MTurk worker who participates in this study. Hence, on top of what you keep for yourself, you will receive what that other person transfers to you. The two MTurk workers you are paired with are not the same.

[a]Loop size 'entire session': 5 sessions recruited with a target of 50, with final session only attracting 9 hits. Data is taken from Wehrli (2018).

Appendix D. Additional figures

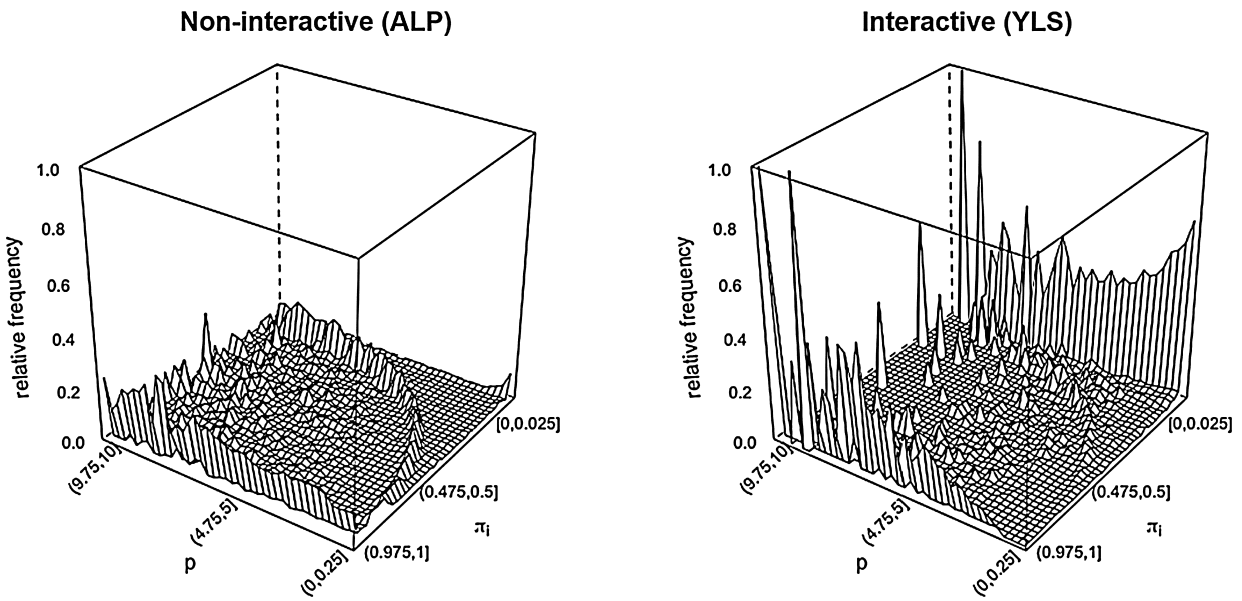

Fig. 5. Giving in Fisman et al. (2015b) – Relative Payment Frequencies for $p \in [0.1, 10]$ for 'elite' vs 'average' samples. *Non-interactive (ALP):* Subjects are representative of the US public and were recruited via the American Life Panel (ALP). The giving patterns resemble those from many previous studies. Many intermediate payments above effective equal-splitting are observed, almost no zero payments, and two spikes at full-transfer and effective equal-splitting (Rawlsian: $p\pi_i = 1 - \pi_i$). *Interactive (YLS):* Subjects were students at Yale Law School (YLS). The giving pattern is extremal and quite unusual compared with previous studies. Fewer intermediate payments above and below effective equal-splitting are observed, less mass at effective equal-splitting, and two spikes at zero and full payments. The mass at zero is decreasing in p, while the mass at everything is increasing in p (note that this latter mass is particularly high compared with the vast majority of prior studies, particularly for high p). *Source:* authors' own figures based on data obtained from Fisman et al. (2015b) and from the ALP, as per Science's data policy.

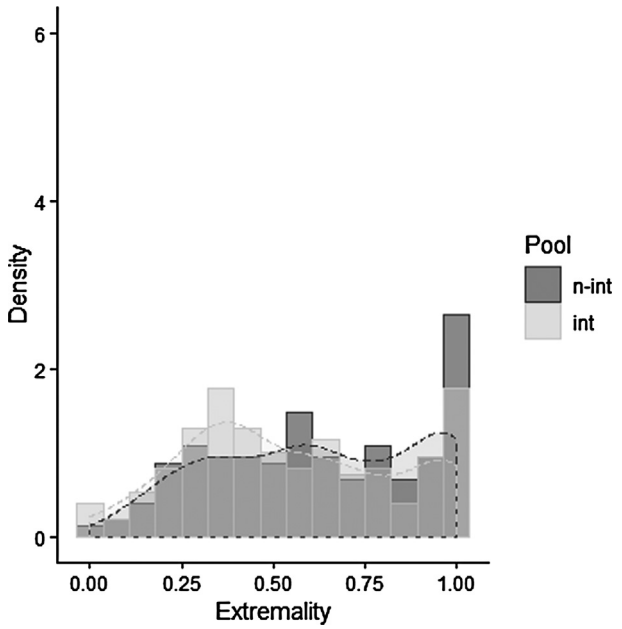

Fig. 6. Extremality of decisions for the interactive and non-interactive protocols. We use the same extremality index as in Fig. 3 (with equal-splitting at 0, and zero-/full-giving at 1).

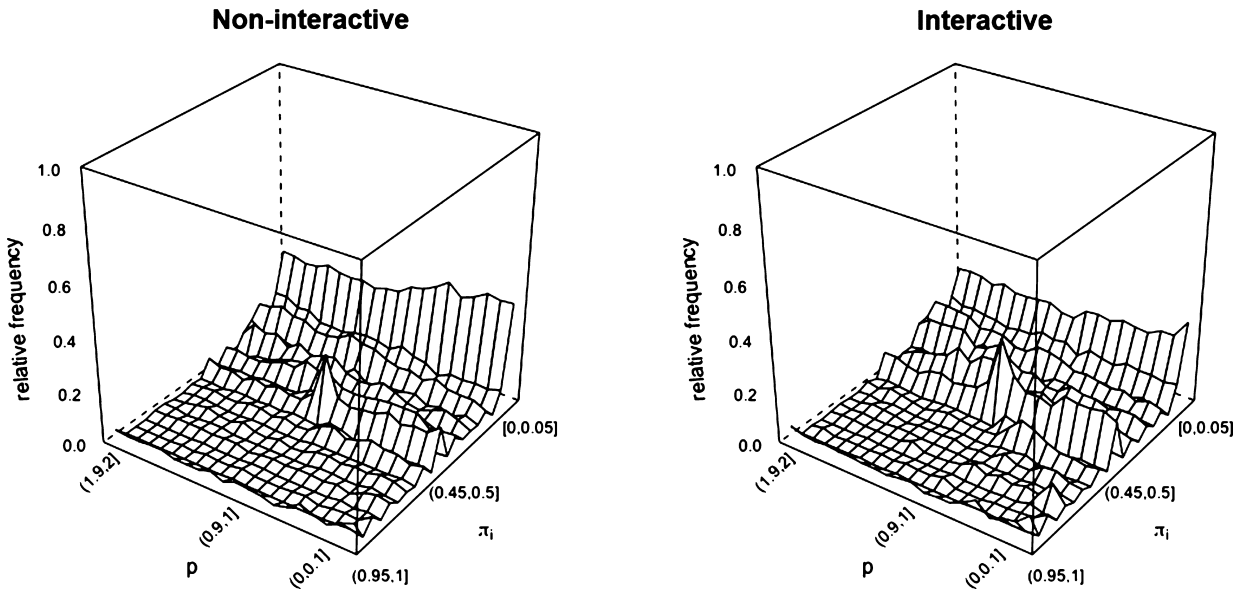

Fig. 7. Giving in our Experiment – Relative Payment Frequencies for $p \in [0.1, 2]$ (note that this range differs from the one in Fig. 5). *Non-interactive:* We observe many intermediate payments, mostly below half, and two spikes at zero-transfer and equal-splitting. Giving decreases in p. *Interactive:* We observe fewer zero-payments, more equal-splitting. The mass at zero is decreasing in p.

Appendix E. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.geb.2019.10.004.

References

Afriat, S.N., 1967. The construction of utility functions from expenditure data. Int. Econ. Rev. 8 (1), 67–77.
Afriat, S.N., 1972. Efficiency estimation of production functions. Int. Econ. Rev. 13 (3), 568–598.
Andreoni, J., 1989. Giving with impure altruism: applications to charity and Ricardian equivalence. J. Polit. Econ. 97 (6), 1447–1458.
Andreoni, J., 1990. Impure altruism and donations to public goods: a theory of warm-glow giving. Econ. J. 100 (401), 464–477.
Andreoni, J., Miller, J., 2002. Giving according to GARP: an experimental test of the consistency of preferences for altruism. Econometrica 70 (2), 737–753.
Andreoni, J., Vesterlund, L., 2001. Which is the fair sex? Gender differences in altruism. Q. J. Econ. 116 (1), 293–312.
Ashraf, N., Bohnet, I., Piankov, N., 2006. Decomposing trust and trustworthiness. Exp. Econ. 9 (3), 193–208.
Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., Sugden, R., 2010. Experimental Economics: Rethinking the Rules. Princeton University Press.
Binmore, K., 2015. Rationality. In: Handbook of Game Theory with Economic Applications, vol. 4. Elsevier, pp. 1–26. Chapter 1.
Bolton, G.E., Ockenfels, A., 2000. ERC: a theory of equity, reciprocity, and competition. Am. Econ. Rev. 90 (1), 166–193.
Camerer, C.F., Ho, T.-H., Chong, J.-K., 2004. A cognitive hierarchy model of games. Q. J. Econ. 119 (3), 861.
Cameron, L., Erkal, N., Gangadharan, L., Meng, X., 2013. Little emperors: behavioral impacts of China's one-child policy. Science 339 (6122), 953–957.
Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. Q. J. Econ. 117 (3), 817–869.
Chaudhuri, A., 2011. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. Exp. Econ. 14 (1), 47–83.
Chen, Y.-C., Van Long, N., Luo, X., 2007. Iterated strict dominance in general games. Games Econ. Behav. 61 (2), 299–315.
Choi, S., Fisman, R., Gale, D.M., Kariv, S., 2007. Revealing preferences graphically: an old method gets a new tool kit. Am. Econ. Rev. 97 (2), 153–158.
Cousineau, D., et al., 2005. Confidence intervals in within-subject designs: a simpler solution to Loftus and Masson's method. Tutor. Quant. Methods Psychol. 1 (1), 42–45.
Crawford, V.P., Costa-Gomes, M.A., Iriberri, N., 2013. Structural models of nonequilibrium strategic thinking: theory, evidence, and applications. J. Econ. Lit. 51 (1), 5–62.
Dufwenberg, M., Heidhues, P., Kirchsteiger, G., Riedel, F., Sobel, J., 2011. Other-regarding preferences in general equilibrium. Rev. Econ. Stud. 78 (2), 613–639.
Engel, C., 2011. Dictator games: a meta study. Exp. Econ. 14 (4), 583–610.
Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Q. J. Econ. 114 (3), 817–868.
Fisman, R., Jakiela, P., Kariv, S., 2015a. How did distributional preferences change during the great recession? J. Public Econ. 128, 84–95.
Fisman, R., Jakiela, P., Kariv, S., 2017. Distributional preferences and political behavior. J. Public Econ. 155, 1–10.
Fisman, R., Jakiela, P., Kariv, S., Markovits, D., 2015b. The distributional preferences of an elite. Science 349 (6254).
Fisman, R., Kariv, S., Markovits, D., 2007. Individual preferences for giving. Am. Econ. Rev. 97 (5), 1858–1876.
Fisman, R., Kariv, S., Markovits, D., 2009. Exposure to Ideology and Distributional Preferences. Working paper.
Forsythe, R., Horowitz, J.L., Savin, N.E., Sefton, M., 1994. Fairness in simple bargaining experiments. Games Econ. Behav. 6 (3), 347–369.
Gigerenzer, G., Selten, R., 2001. Bounded Rationality: The Adaptive Toolbox. MIT Press.
Grech, P.D., 2019. Give and let give: alternative mechanisms based on voluntary contributions. Games 10 (2), 21.
Grech, P.D., Nax, H.H., 2017. Interactive vs. non-interactive dictator games. In: Proceedings of the International Conference on Group Decision and Negotiation, vol. 17. In: Hohenheim Discussion Papers in Business, Economics and Social Sciences, pp. 309–313. Chapter Extended Abstract.
Greiff, M., Ackermann, K.A., Murphy, R.O., 2018. Playing a game or making a decision? Methodological issues in the measurement of distributional preferences. Games 9 (4), 80.
Iriberri, N., Rey-Biel, P., 2011. The role of role uncertainty in modified dictator games. Exp. Econ. 14 (2), 160–180.
Jakiela, P., 2013. Equity vs. efficiency vs. self-interest: on the use of dictator games to measure distributional preferences. Exp. Econ. 16 (2), 208–221.

Kahneman, D., Knetsch, J.L., Thaler, R.H., 1986. Fairness and the assumptions of economics. J. Bus. 59 (4), S285–S300.
Korenok, O., Millner, E.L., Razzolini, L., 2013. Impure altruism in dictators' giving. J. Public Econ. 97, 1–8.
Krupka, E.L., Weber, R.A., 2013. Identifying social norms using coordination games: why does dictator game sharing vary? J. Eur. Econ. Assoc. 11 (3), 495–524.
Levitt, S.D., List, J.A., 2007. What do laboratory experiments measuring social preferences reveal about the real world? J. Econ. Perspect. 21 (2), 153–174.
Li, J., Dow, W.H., Kariv, S., 2017. Social preferences of future physicians. Proc. Natl. Acad. Sci. 114 (48), E10291–E10300.
Liebrand, W.B.G., 1984. The effect of social motives, communication and group size on behaviour in an n-person multi-stage mixed-motive game. Eur. J. Soc. Psychol. 14 (3), 239–264.
Mäs, M., Nax, H.H., 2016. A behavioral study of 'noise' in coordination games. J. Econ. Theory 162, 195–208.
Morey, R.D., et al., 2008. Confidence intervals from normalized data: a correction to Cousineau (2005). Reason 4 (2), 61–64.
Murphy, R.O., Ackermann, K.A., Handgraaf, M.J.J., 2011. Measuring social value orientation. Judgm. Decis. Mak. 6 (8), 771–781.
Nax, H.H., Murphy, R.O., Ackermann, K.A., 2015. Interactive preferences. Econ. Lett. 135, 133–136.
Oechssler, J., 2010. Searching beyond the lamppost: let's focus on economically relevant questions. J. Econ. Behav. Organ. 73 (1), 65–67.
Polisson, M., Quah, J., 2013. Revealed preference in a discrete consumption space. Am. Econ. J. Microecon. 5 (1), 28–34.
Polisson, M., Renou, L., 2016. Afriat's theorem and Samuelson's 'Eternal Darkness'. J. Math. Econ. 65, 36–40.
Samuelson, P.A., 1938. A note on the pure theory of consumer's behaviour. Economica 5 (17), 61–71.
Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol. Sci. 22 (11), 1359–1366.
Sobel, J., 2005. Interdependent preferences and reciprocity. J. Econ. Lit. 43 (2), 392–436.
Solow, R.M., 1956. A contribution to the theory of economic growth. Q. J. Econ. 70 (1), 65–94.
Wehrli, S., 2018. iDictator: Interactive dictator games with varying loop sizes. Mimeo, osf.io/zf8c5/.